GESELLSCHAFT
FÜR INFORMATIK



Daniel Demmler, Daniel Krupka, Hannes Federrath (Hrsg.)

INFORMATIK 2022

Informatik in den Naturwissenschaften

26. – 30. September 2022
Hamburg

Gesellschaft für Informatik e.V. (GI)

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

Volume P-326
ISBN 978-3-88579-720-3
ISSN 1617-5468

Volume Editors
Daniel Demmler
Universität Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany
daniel.demmler@uni-hamburg.de
Daniel Krupka
Gesellschaft für Informatik, Anna-Louisa-Karsch-Straße 2, 10178 Berlin, Germany
daniel.krupka@gi.de
Hannes Federrath
Universität Hamburg, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany
hannes.federrath@uni-hamburg.de

Series Editorial Board
Andreas Oberweis, KIT Karlsruhe,
(Chairman, andreas.oberweis@kit.edu)
Torsten Brinda, Universität Duisburg-Essen, Germany
Dieter Fellner, Technische Universität Darmstadt, Germany
Ulrich Flegel, Infineon, Germany
Ulrich Frank, Universität Duisburg-Essen, Germany
Michael Goedicke, Universität Duisburg-Essen, Germany
Ralf Hofestädt, Universität Bielefeld, Germany
Wolfgang Karl, KIT Karlsruhe, Germany
Michael Koch, Universität der Bundeswehr München, Germany
Peter Sanders, Karlsruher Institut für Technologie (KIT), Germany
Andreas Thor, HFT Leipzig, Germany
Ingo Timm, Universität Trier, Germany
Karin Vosseberg, Hochschule Bremerhaven, Germany
Maria Wimmer, Universität Koblenz-Landau, Germany

Dissertations: Steffen Hölldobler, Technische Universität Dresden, Germany
Thematics: Agnes Koschmider, Universität Kiel, Germany
Seminars: Judith Michael, RWTH Aachen, Germany

© Gesellschaft für Informatik, Bonn 2022
printed by Köllen Druck+Verlag GmbH, Bonn



This book is licensed under a Creative Commons BY-SA 4.0 licence.

Vorwort

Die INFORMATIK 2022 ist die offizielle Jahrestagung der Gesellschaft für Informatik e.V. (GI), der größten Vereinigung der Informatikerinnen und Informatiker im deutschsprachigen Raum und wird jährlich an wechselnden Orten veranstaltet. Die 52. Jahrestagung INFORMATIK 2022 findet vom 26. bis 30.09.2022 in Hamburg als Präsenzveranstaltung statt. Nach zwei aufgrund der pandemischen Situation weitgehend virtuell durchgeführten Jahrestagungen freuen wir uns auf die Begegnungen in den Tagungsräumen der Universität Hamburg.

Das Leitthema der INFORMATIK 2022 lautet „Informatik in den Naturwissenschaften". Dieses Thema fokussiert auf die Synergien der verschiedenen Naturwissenschaftsdisziplinen mit der Informatik. Naturwissenschaften beobachten und ergründen die Natur durch Empirie und Experimente, um sie verstehbar und nutzbar zu machen. Die Informatik nimmt dabei eine immer zentralere Rolle ein, indem sie datengetrieben komplexe Analysen und Vorhersagen ermöglicht und große Datenmengen verarbeitet, verwaltet und visualisiert. Einerseits bedienen sich die Naturwissenschaften dabei der Informatik als Werkzeug, andererseits ist die Natur in den verschiedenen Bereichen der Informatik ein Vorbild und Quelle für Inspiration.

Die fast 150 schriftlichen Beiträge aus fast 30 Workshops und Tutorials der INFORMATIK 2022 sind in diesem Tagungsband zusammengefasst.

Neben den Beiträgen zum Schwerpunktthema „Informatik in den Naturwissenschaften" gibt die INFORMATIK 2022 einen Einblick zu weiteren Themen, die aktuell besondere Aufmerksamkeit in Wissenschaft, Gesellschaft, Verwaltung, Wirtschaft und Praxis erfahren. Dies sind z. B. Web Science, Data Science, Internet of Everything, Sichere und zuverlässige Systeme, Künstliche Intelligenz und Maschinelles Lernen, Autonomes Fahren und Mobilität, Quantencomputing, Digital Health, Green Coding, Robotik, Industrie 4.0, Software Engineering, Computerethik sowie Informatik-Didaktik.

Auf der INFORMATIK 2022 bringen wir Gesprächspartnerinnen und -partner aus Wissenschaft, Politik und Wirtschaft zusammen, beleuchten gesellschaftliche und politische Zusammenhänge und suchen den Dialog mit allen fachlich Interessierten und der Öffentlichkeit.

Im Rahmen der INFORMATIK 2022 wird – wie auch in den Vorjahren – mit der EnviroInfo 2022 auch die 36. Internationale Konferenz zu Umweltinformations- und Kommunikationstechnologie veranstaltet. Die englischsprachige EnviroInfo widmet sich in Special Tracks etwa den Themen Green Data Centres, Environmental Informatics education, Economizing Nature, Circular Economy und Industrial Symbiosys. Auch hier stehen zudem KI-Themen in Fokus.

Auf der INFORMATIK 2022 findet auch die 12. SKILL- Studierendenkonferenz Informatik statt. Die Studierendenkonferenz der Gesellschaft für Informatik richtet sich an alle Studentinnen und Studenten der Informatik sowie angrenzender Disziplinen. Im Mittelpunkt der Konferenz stehen studentische Arbeiten und Projekte. Bitte beachten Sie, dass Einreichungen von Doktorandinnen und Doktoranden sowie Einreichungen mit

Ko-Autorenschaften von Professorinnen und Professoren nicht zugelassen werden können. Angenommene Beiträge werden in einem eigenen Tagungsband publiziert und auf der SKILL am 29 und 30. September 2022 im Rahmen der INFORMATIK 2022 präsentiert. Für herausragende Beiträge werden Preise ausgelobt.

Neben den Preisen der SKILL werden im Rahmen der INFORMATIK 2022 auch wieder eine Vielzahl an weitere Preisen verliehen: der GI-Dissertationspreis wird mit Unterstützung der der Österreichischen Computergesellschaft (OCG) und der Schweizer Informatik Gesellschaft (SI) vergeben, der Umweltinformatikpreis des Fachausschuss Umweltinformatik sowie der Preis des Fakultätentages. Im Rahmen des festlichen Konferenzdinners wird zudem ein neuer Jahrgang an Fellows und Junior Fellows benannt sowie die Klaus-Tschira-Medaille für besondere Verdienste um die Nutzung und Weiterentwicklung informatischer Methoden in unterschiedlichen Anwendungsgebieten vergeben.

Erstmals wird in diesem Jahr auch der von Helmut und Heide Balzert gestiftete Preis für einen herausragenden Beitrag zur Vermittlung von Inhalten der Informatik und zur Ausbildung/Didaktik der Informatik (IAD) vergeben. Ziel des von der Gesellschaft für Informatik e.V. (GI) in Kooperation mit der Österreichischen Computergesellschaft (OCG) und der Schweizer Informatik Gesellschaft (SI) verliehenen Preises ist es, die Aufmerksamkeit für die Bedeutung der „digitalen Didaktik“ bei der Vermittlung von komplexen Sachverhalten in der Informatik zu fördern und zu unterstützen. Der Preis soll die Entwicklung und Erprobung innovativer Konzepte für digitale Lehr- und Lernformen fördern und ist mit einem Preisgeld in Höhe von 10.000 € dotiert.

Den Leserinnen und Lesern dieses Tagungsbandes wünschen wir nun viel Spaß bei der Lektüre.

Hamburg im September 2022

Daniel Demmler, Daniel Krupka und Hannes Federrath

Sponsoren

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

Google Deutschland GmbH



adesso SE



genua GmbH



Hamburger Informatik Technologie-Center (HITeC) e.V.



SAP SE



Capgemini



ADDITIVE - Soft- und Hardware für Technik und Wissenschaft GmbH



Walter de Gruyter GmbH



Springer Vieweg



IAV GmbH



DESY



Tagungsleitung

Gesamtleitung:	Hannes Federrath, Universität Hamburg
	Daniel Krupka, Gesellschaft für Informatik
Leitung enviroInfo2022:	Volker Wohlgemuth, htw Berlin
Leitung SKILL 2022:	Michael Becker, Universität Leipzig
	Judith Michael, RWTH Aachen
	Thomas Riechert, HTWK Leipzig
	Ludger Porada, Gesellschaft für Informatik
Koordination Hauptprogramm:	Maike Klein, Gesellschaft für Informatik
Koordination Workshops:	Daniel Demmler, Universität Hamburg
Tagungsband:	Christian Burkert, Universität Hamburg

Organisationsteam

Daniel Demmler	Universität Hamburg
Britta Skulima	Universität Hamburg
Cornelia Winter	Gesellschaft für Informatik
Maike Klein	Gesellschaft für Informatik
Alexander Scheibe	Gesellschaft für Informatik
Maximilian Weinl	Gesellschaft für Informatik
Frithjof Nagel	Gesellschaft für Informatik
Elisabeth Schauermann	Gesellschaft für Informatik
Nikolas Becker	Gesellschaft für Informatik
Anna-Sarah Lieckfeld	Gesellschaft für Informatik

Programmbeirat

Hannes Federrath	Universität Hamburg
Stefan Kast	TU Dortmund
Richard Lenz	Friedrich-Alexander-Universität Erlangen-Nürnberg
Astrid Nieße	Universität Oldenburg
Matthias Rarey	Universität Hamburg
Norbert Ritter	Universität Hamburg
Heidi Seibold	selbständig
Volker Wohlgemuth	htw Berlin
Fereshta Yazdani	Lufthansa Industry Solutions

Workshopchairs

Andreas Abecker	disy Informationssysteme GmbH
Tobias Bahr	Universität Stuttgart
Franziska Boenisch	Fraunhofer AISEC
Carsten Brockman	SAP Deutschland
Christina B. Class	Ernst-Abbe-Hochschule Jena
Christian Czarnecki	FH Aachen
Jan de Meer	smartspacelab.eu
Anna-Katharina Dhungel	Universität Lübeck
Daniel Eberz-Eder	DLR-RNH
Marlene Eisenträger	VDI/VDE-IT
Felix Gessert	Universität Hamburg
Matthias Goeken	Hochschule der Deutschen Bundesbank
Ulrich John	International University of Applied Sciences
Reinhard Kahle	Universität Tübingen
Karolin Eva Kappler	FernUniversität Hagen
Natalja Kleiner	FZI Forschungszentrum Informatik
Alexander Kosik	selbständig
Edna Kropp	LivePerson
Jörg Lässig	Hochschule Zittau/Görlitz
Viktor Leis	Friedrich-Alexander-Universität Erlangen-Nürnberg
Kerstin Lenk	Justus-Liebig-Universität Gießen
Verena Majuntke	HTW Berlin
Wolfgang Mauerer	Technische Hochschule Regensburg
Vera G. Meister	TH Brandenburg
Mario Pesch	Universität Münster
Roman Povalej	Polizeiakademie Niedersachsen
Steffi Rudel	Universität der Bundeswehr München
Tim Ruhe	TU Dortmund
Tim Philipp Schäfers	FHDW Paderborn
Annika Selzer	Universität Bremen
Thorsten Schoormann	Universität Hildesheim
Christoph Sorge	Universität des Saarlandes
Jan Wegener	
Axel Wiepke	Universität Potsdam
Wolfram Wingerath	Universität Oldenburg

Inhaltsverzeichnis

Workshops

International Workshop On Digital Forensics (IWDF)

Roman Povalej
2nd International Workshop on Digital Forensics – an Interexchange of Law Enforcement and Science 33

Jenny Felser, Jian Xi, Christoph Demus, Dirk Labudde, Michael Spranger
Recommendation of Query Terms for Colloquial Texts in Forensic Text Analysis 35

Maximilian Eisenhut, Wilfried Honekamp
Potential analysis for the detection of attacks on wireless networks using the Wireless Intrusion Detection System Nzyme 49

Viktor Garske, Andreas Noack
Revisiting the privacy of censored credentials 59

Martin Morgenstern, Johannes Fähndrich, Wilfried Honekamp
Ontology in the Digital Forensics Domain: A Scoping Review 71

Alessia Cornaggia-Urrigshardt, Nikita Jarocky, Frank Kurth, Sebastian Urrigshardt, Kevin Wilkinghoff
SCALA-Speech: An Interactive System for Finding and Analyzing Speech Content in Audio Data 81

Florian Heinke, Marie-Luise Heuschkel, Dirk Labudde
A frequentist estimation of duplicate probability as a baseline for person identification from image and video material using anthropometric measurements 91

Daniel Spiekermann, Jörg Keller
Challenges of Network Traffic Classification Using Deep Learning in Virtual Networks 99

Svenja Preuß, Dirk Labudde
A pipeline for analysing image and video material in a forensic context with intelligent systems 109

Daniel Vogel, Markus Krämer
Collecting Identifying Data for Re-Identification of Mobile Devices carried at a Crime Scene using Wi-Fi Routers 119

Mehr Frauen in die IT – Wie bekommen wir das hin?

Janki Dodiya, Cornelia Heinisch, Claudia Hess, Sibylle Kunz, Sandra Rebholz, Inga Schlömer, Adrienne Steffen, Silke Vaas
Yes she can! How to encourage girls to choose a career path in computer science 131

Nadine Dittert, Ira Diethelm
Alex wird Informatikerin - Eine Geschichte von Studierenden für Jugendliche zur Berufsinformation 135

Carolin Neumann, Clarissa Sabrina Arlinghaus
STEM Girls – Eine Workshopreihe zur Förderung von Frauen und nicht-binären Personen in STEM-Berufen 139

Nicola Marsden, Karen Holtzblatt
Making Remote Work for Women in Tech 151

3. Interdisciplinary Privacy & Security at Large Workshop (Privacy&Security@Large)

Annika Selzer, Ingo J. Timm
3. Interdisciplinary Privacy & Security at Large Workshop 159

Marian Eleks, Jonas Rebstadt, Philipp Fukas, Oliver Thomas
Learning without Looking: Similarity Preserving Hashing and Its Potential for Machine Learning in Privacy Critical Domains 161

Sarah Stummer
Issues of Verifying Anonymity: An Overview 179

K. Valerie Carl
The status-quo of companies' data privacy and security communication: An ethical evaluation and future paths 195

Niklas Sänger, Sebastian Abeck
Authentication and Authorization in Microservice-Based Applications . . 207

Maja Schneider, Lukas Gehrke, Peter Christen, Erhard Rahm
D-TOUR: Detour-based point of interest detection in privacy-sensitive trajectories 219

Janine Schleper, Matthias Kohn, Paulina Jo Pesch, Ulrich Waldmann, Thomas Kunz
Messung der Datenminimierung für den Beschäftigtendatenschutz am Beispiel von Standortdaten 231

Linda Schreiber
Rechtliche Implikationen von Cybersicherheitsvorfällen in der kommunalen Verwaltung 237

Künstliche Intelligenz für kleine und mittlere Unternehmen (KI-KMU 2022)

Maximilian Lowin, Jendrik Seip
AI-based Online Vacancies - Trends and Differences between SMEs and Larger Enterprises 247

Laslo Welz, Helmut Beckmann
Anwendung von Machine Learning bei der datengetriebenen Prozessanalyse - Eine State-of-the-Art Literaturanalyse 259

Janos Borst, Christofer Meinecke, Daniel Wiegreffe, Andreas Niekler
A Demonstration System towards NLP and Knowledge Driven Data Platforms for Civil Engineering 271

Henrik Kortum, Jonas Rebstadt, Tula Böschen, Pascal Meier, Oliver Thomas
Towards the Operationalization of Trustworthy AI: Integrating the EU Assessment List into a Procedure Model for the Development and Operation of AI-Systems 283

Walter Laurito, Jacqueline Höllig, Jonas Lachowitzer, Steffen Thoma, Matthias Budde, Patrick Philipp
AIDA-Vis – Automatic Data Visualization with Human Preferences 301

Patrick Berger, Jörg von Garrel
Nutzenbasierte Preisgestaltung kooperativer KI-basierter Software in produzierenden Unternehmen – Eine empirische Untersuchung KI-basierter Geschäftsmodelle 307

Nachhaltige Wertschöpfungssysteme (NaWerSys) II

Thorsten Schoormann, Friedemann Kammler, Paul Christoph Gembarski, Simon Hagen
2. Workshop: Nachhaltige Wertschöpfungssysteme (NaWerSys) 325

Julia Christina Schweihoff, Ilka Jussen, Maleen Stachon, Frederik Möller
Design Options for Data-Driven Business Models in Data-Ecosystems . . 331

Jonas Brinker, Friedemann Kammler, Oliver Thomas
Service Tailoring: Ein Konfigurationsmechanismus für individualisierte After-Sales-Services im Maschinen- und Anlagenbau 347

Henrik Kortum, Tobias Kohl, Dominik Hubertus, Oliver Hinz, Oliver Thomas
A Platform Framework for the Adoption and Operation of ML-based Smart Services in the Data Ecosystem of Smart Living 361

Ilka Jussen, Julia Christina Schweihoff, Maleen Stachon, Frederik Möller
Designing a Data Sharing Tool Kit - Showing companies how to start using Data Sharing 379

Jessica Bollenbach, Stefan Neubig, Andreas Hein, Robert Keller, Helmut Krcmar
Using Machine Learning to Predict POI Occupancy to Reduce Overcrowding 393

Dimitri Petrik, Stefan Trieflinger, Felix Schönhofen
Aligning Value Creation in Ecosystems through Roadmapping 409

Konzeptionelle Herausforderungen für die KI: Naturwissenschaften

Klaus Mainzer
Konzeptionelle Grundlagen der Künstlichen Intelligenz: KI und Naturwissenschaften 427

Workshop on Machine Learning for Astroparticle Physics and Astronomy (ml.astro)

Mirko Bunse
Unification of Algorithms for Quantification and Unfolding 459

Janis Kummer, Lennart Rustige, Florian Griese, Kerstin Borras, Marcus Brüggen, Patrick L. S. Connor, Frank Gaede, Gregor Kasieczka, Peter Schleper
Radio Galaxy Classification with wGAN-Supported Augmentation 469

Dmitry Malyshev, Aakash Bhat
Towards probabilistic multiclass classification of gamma-ray sources . . . 479

Shreyas Kalvankar, Hrushikesh Pandit, Pranav Parwate, Atharva Patil, Snehal Kamalapur
Astronomical Image Colorization and Up-scaling with Conditional Generative Adversarial Networks 489

Pranav Sampathkumar, Augusto Antonio Alves Junior, Tanguy Pierog, Ralf Ulrich
Sequential networks for cosmic ray simulations 499

Jigar Bhanderi, Stefan Funk, Dmitry Malyshev, Naomi Vogel, Andreas Zmija
Calculation of the Photon Flux in a Photo-Multiplier Tube with Deep Learning 507

Recht und Technik: Datenschutz im Diskurs (RuT)

Rüdiger Grimm, Gerrit Hornung, Christoph Sorge, Indra Spiecker genannt Döhmann
Recht und Technik – Datenschutz im Diskurs 519

Nicolas Boltz, Leonie Sterz, Christopher Gerking, Oliver Raabe
A Model-Based Framework for Simplified Collaboration of Legal and Software Experts in Data Protection Assessments 521

Berna Orak, Erik Krempel, Arno Appenzeller
Datenschutzkonforme Weitergabe von Versichertendaten aus dem Forschungsdatenzentrum 533

Marc Widemann
Bereitstellung von Mobilitätsdaten – Ein Problem des Datenschutzes? . . 551

Sandra Wittmer, Florian Platzer
Zulässigkeit von Open Source-Ermittlungen zur Strafverfolgung im Darknet 569

Thomas Kunz, Ulrich Waldmann
ML-basierte Klassifizierung von E-Mails für die datenschutzkonforme Löschung und Archivierung 589

Dirk Müllmann, Maxime Veit, Melanie Volkamer
Technische und Rechtliche Auseinandersetzung mit Weiterleitungs-URLs in E-Mails aus Security-Gründen 601

Tim Wambach
Sicherheits- und Datenschutzanalyse der E-Mail-Server von Krankenhäusern und Kliniken 617

Konstantin Knorr, David Müller
Automatisierte Überprüfung von Webauftritten auf Fremdinhalte 639

Inna Vogel, Tahireh Setz, Jeong-Eun Choi, Martin Steinebach
Natural Language Processing (NLP) und der Datenschutz - Chancen und Risiken für den Schutz der Privatheit 651

Simone Salemi, Bianca Steffes
Deepfakes im VideoIdent-Verfahren: (fehlende) Straf- und zivilrechtliche Konsequenzen für Täter 665

Informatik & Gesellschaft - Zukunft gestalten / IUG2022

Kerstin Raudonat, Monika Pröbster, Kai Schmieder, Simone Martinetz, Nicola Marsden
Where bias can creep in - Gendersensibilität beim Einsatz von KI-Technologien im Kontext beruflicher Weiterbildung im Forschungsprojekt KIRA 683

Karolin Eva Kappler, Nicole Wolf, Tanja Kopp-Malek, Dominik Pataky, Christiane Bodammer, Virginia Holness, Christoph Reithmair
Digitalisierung – Brandbeschleuniger oder Helfer eines nachhaltigen Wandels? Nachhaltigkeitskriterien für digitale Werkzeuge und Projekte . . 693

Andy Börner, Verena Traubinger, Philipp Stiens, Maximilian Eibl
Das Universal Paradox: Im Spannungsfeld zwischen Universal Design und User Experience in einer Indoor-Navigations App für Menschen mit Einschränkungen 703

Nils Pancratz, Matthias Matzner
Die ethische Position von Informatikstudierenden bezüglich der Entwicklung von Informatiksystemen 711

Karolin Eva Kappler, Katharina Ebner, Stefan Smolnik
We are smart, smarter, the smartest?! - Smart systems beyond optimization 717

Ole Lemke, Jürgen Altmann, Martin Degeling
Crisis escalation from autonomous weapon systems: agent-based modeling for military-technology assessment 723

Sercan Sever
Informatisches Gestalten unter Bedingungen von Gegenwartskomplexität und Zukunftsunsicherheit. Plädoyer für eine Verständigung in und mit Gesellschaft 729

Gabriele Kunau, Andrea Kienle
Die ethischen Leitlinien der GI als Rahmen für Informatik und Gesellschaft in der Bachelorausbildung 735

Claudia Müller-Birn, Lars Sipos
Human-Centered Data Science – Etablierung einer kritisch-reflexiven Praxis bei der Entwicklung von datengetriebener Software 745

Hermann Diebel-Fischer, Lutz Hellmig
Ethische Fragen und informatische Modelle – eine symbiotische Beziehung in der informatischen Bildung 755

12. Betriebliche Umweltinformationssysteme (BUIS-Tage 2022)

Harish Moturu, Jorge Marx Gómez, Johannes Schering
A New Approach to Determine Cycling KPIs in the Context of Behavioral Data 763

Bettina Steden, Jorge Marx Gómez
Elektromobilität in Unternehmen - Schwerpunkte und Methoden der Forschung 771

Johannes Schering, Jorge Marx Gómez, Lena Büsselmann, Federico Alfaro, Jan Stüven
Potentials of Bicycle Infrastructure Data Lakes to Support Cycling Quality Assessment 783

Herbert Brüning, Jorge Marx Gómez, Alexander Papke, Sreedhar Kokkarachedu
Mobilitätswende @ Lebensqualität (MobiLe) – Die Entwicklung eines qualitativen Modells zur Vermittlung wichtiger Wirkungszusammenhänge im komplexen System Verkehr 795

Jantje Halberstadt, Arne Ortland, Anne-Kathrin Schwab, Andrii Besieda, Martin Kraft
Digitalisierung und Innovation in der Landwirtschaft 807

Jana Götze, Wilko Hinz, Mario Päßler, Axel Hiller
Digitale Wandlungsprozesse: 30 Jahre Bergbausanierung im Spiegel der Umweltdatenerhebung 823

Marius Wybrands, Sven von Höveling, Marco Muras, Heinrich Töpfer, Jorge Marx Gómez
Prototypische Implementierung eines Digital Twin und Predictive Maintenance Methoden am Beispiel des Wasserwerks Langeoog 831

Ali Akyol, Andreas Solsbach, Jorge Marx Gómez
Entwicklung eines bedarfsgerechten Umweltinformationsportals mit Bürgerbeteiligung 843

Heiko Thimm, Vanessa Schmidt
Low-Code Basierte Entwicklung eines Chatbots für Betriebliches Compliance Management im Umweltbereich 855

Wirtschaftsinformatik an Hochschulen Angewandter Wissenschaften

Christian Czarnecki, Vera G. Meister, Verena Majuntke
Wirtschaftsinformatik an Hochschulen Angewandter Wissenschaften . . . 869

Daniel Szafarski, Helmut Beckmann
Einsatz von Machine Learning im Innovationsmanagement 871

Pablo Stockhausen, Andreas Johannsen, Robert Maurer
Konzeption eines Machine-Learnings-Verfahrens zum Lösen von Green Vehicle Routing Problemen 885

Mazlum Copurkuyu, Thomas Barton
Extraktion und Analyse von Schlüsselwörtern in einer Literaturrecherche zu Quantum Computing 897

Frank Neumann, Juan Carlos Rodríguez-del-Pino, Sebastian Homer
Concept, Implementation and Evaluation of a Virtual Learning Environment for Acquiring Competences in Android App Development . . 913

Resilient Smart Farming Lab (RSFLab)

Christian Reuter, Daniel Eberz-Eder, Franz Kuntke, Matthias Trapp
RSF-Lab'22: Resilient Smart Farming Laboratory: Für eine widerstandsfähige und intelligente Landwirtschaft 931

Paul Schulze, Frank Fuchs-Kittowski, Tim Hafemeister, Alexander Urban, Mario Berndl, Christian Simmet, Martin Schulze
Development of a measuring system for monitoring transport of boar semen from artificial insemination centers to sow farms 935

Eike Gassen, Patrick Wolf, Karsten Berns
A modular control architecture for safe and robust robot operation and inspection in steep slope vineyards 947

Christian Reuter, Franz Kuntke, Matthias Trapp, Christian Wied, Gerwin Brill, Georg Müller, Enno Steinbrink, Jonas Franken, Daniel Eberz-Eder, Wolfgang Schneider
AgriRegio: Infrastruktur zur Förderung von digitaler Resilienz und Klimaresilienz im ländlichen Raum am Beispiel der Pilotregion Nahe-Donnersberg 961

Lea Wintz, Paul Strerath
Aufbau und Umsetzung eines informellen Netzwerkes zur Förderung der Digitalisierung in der Landwirtschaft in Rheinland-Pfalz zur Implementierung einer Wissenstransfer-Plattform 973

Laura Buhleier, Patrick Gantner, Tobias Frey, Michael Bors, Marc-André Kaufhold, Christian Reuter
Effizienz und Nachhaltigkeit durch Green-IT: ein systematischer Literaturüberblick im Kontext der Klimakrise 995

Datenqualität und Qualitätsmetriken in der Datenwirtschaft (DQ)

Gesa Marie Götte, Bonito Steffen Thielert, Andreas Herzog
Application-specific quality metrics for the assessment of data for deep learning from large datasets 1015

Agbodzea Pascal Ahiagble, Hannah Stein
Approaches for Automated Data Quality Analysis: Syntactic and Semantic Assessment 1023

Dusan Dokic, Hannah Stein
Towards a data quality index for data valuation in the data economy . . . 1037

Philipp Schlunder
Teilautomatisierte Datenqualitätsbewertung und Fehlerkorrektur zur Senkung der Einstiegshürde von Datenanalysemethoden 1055

Isabell Viedt, Jonathan Mädler, Valentin Khaydarov, Leon Urbas
Prescriptive and descriptive quality metrics for the quality assessment of operational data 1061

Agile Methoden im Informatikunterricht

Tobias Bahr, Bernd Zinn
Workshop zum Einsatz agiler Methoden im Informatikunterricht 1067

Geodaten als open data für die Künstliche Intelligenz

Anna Hoffmann, Reinhard Kahle
Geodaten als Open Data für die Künstliche Intelligenz 1081

QSAP - Quantum Software and Applications

Peter Faber, Helena Liebelt
High-Performance and Quantum Computing for Students 1085

Markus Zajac, Michelle Dang, Uta Störl
Implementation Aspects for Linking Data in Hybrid Quantum Applications 1093

Gergely Gálfi, Tamás Kozsik, Zoltán Zimborás
Programming language for unified computing with classical and quantum bits 1101

GI Quantum Computing Workshop

Kamalika Datta, Abhoy Kole, Indranil Sengupta, Rolf Drechsler
Mapping Quantum Circuits to 2-Dimensional Quantum Architectures . . 1109

Armin Moin, Moharram Challenger, Atta Badii, Stephan Günnemann
Towards Model-Driven Engineering for Quantum AI 1121

Alexander Mandl, Uwe Egly
Implementations for Shor's algorithm for the DLP 1133

Marika Federer, Daniel Müssig, Steve Lenk, Jörg Lässig
Real-world application benchmark for QAOA algorithm for an electromobility use case 1145

Volker Reers, Jörg Lässig
A new Pattern for Quantum Evolutionary Algorithms 1153

Maximilian Selch, Daniel Müssig, Albrecht Hänel, Jörg Lässig, Steffen Ihlenfeldt
Influence Estimation In Multi-Step Process Chains Using Quantum Bayesian Networks 1163

Daniel Müssig, Jörg Lässig
Quantum Computer and Quantum Algorithm Benchmarking 1175

(Agiles) Enterprise Architecture Management in Forschung und Praxis

Gebhard Herget, Eldar Sultanow, Alina Chircu, Johannes Ludsteck, Sebastian Hammer, Christian Koch, Willy Reuter, Matthias Seßler
Infrastructure anomaly detection: A cloud-native architecture at Germany's Federal Employment Agency 1181

Dirk Bamberger, Farid Okhovat, Carsten Brockmann, Martina Sturm
Flexibilität in Cloud ERP Systemen für bessere Unternehmensprozesse . . 1195

Ann-Kathrin Losse, Melanie Gehrke, André Ullrich, Christian Czarnecki, Eldar Sultanow, Carsten Breithaupt, Christian Koch
Entwicklung einer Open-Data-Referenzarchitektur für die Luftfahrtindustrie 1203

Der EmiLe Computer Club (ECC): Informatik an Schulen spannend und interessant vermitteln

Steffi Rudel
Programmieren lernen an einer Montessori-Schule: Der EmiLe Computer Club (ECC) 1213

GI Startup Platform – Community Get Together

Wolfram Wingerath, Viktor Leis, Felix Gessert
GI Startup Platform – Community Get Together 1223

Trustworthy AI in Science and Society

Oussama Bouanani, Franziska Boenisch
The Influence of Training Parameters on Neural Networks' Vulnerability to Membership Inference Attacks 1227

Thora Markert, Fabian Langer, Vasilios Danos
GAFAI: Proposal of a Generalized Audit Framework for AI 1247

Angelie Kraft, Hans-Peter Zorn, Pascal Fecht, Judith Simon, Chris Biemann, Ricardo Usbeck
Measuring Gender Bias in German Language Generation 1257

Karla Markert, Afrae Ahouzi, Pascal Debus
Fairness in Regression – Analysing a Job Candidates Ranking System . . 1275

eGov-FemTech

Mareike Winkler, Sonja Köhne, Miriam Klöpper
'Not all algorithms!' Lessons from the Private Sector on Mitigating Gender Discrimination 1289

Anna-Katharina Dhungel
How gender is understood and analyzed in current egovernment research – a scoping literature review 1305

Hochschule 2032

Michael Goedicke, Ulrike Lucke
Research Data Management in Computer Science - NFDIxCS Approach . 1317

Heike Neuroth, Janine Straka, Carsten Schneemann, Daniela Mertzen
Kooperativer Aufbau und nachhaltiger Betrieb der Landesinitiative für Forschungsdatenmanagement in Brandenburg (FDM-BB) 1329

Christian Riedel, Hendrik Geßner, Anja Seegebrecht, Safial Islam Ayon, Shafayet Hossen Chowdhury, Ralf Engbert, Ulrike Lucke
Including Data Management in Research Culture Increases the Reproducibility of Scientific Results 1341

Mirjam Blümm, Christine Burkart, Maria Chlastak, Katharina Fritsch, Heike Neuroth, Andreas Schieberle, Stefan Schmunk, Stephanie Werner
Wie steht es um die Sichtbarkeit, Akzeptanz und Nutzung von Datenmanagementplänen an Fachhochschulen? 1353

Bert Zulauf, Nina Knipprath
Erfahrungsbericht aus dem Forschungsdatenmanagement in Bezug auf elektronische Laborbücher 1359

Stephan Leible, Max Ludzay
A Snapshot of Essential IT-Related Challenges of Universities: A Literature Analysis 1365

Markus von der Heyde
What's Digital? Factors of perceived Digital Transformation of Higher Education in Germany 1379

Kevin Saukel
Bildungsinformatik als Accelerator innovativer Veränderungsprozesse in der Hochschule der Zukunft 1393

Stefanie Lemcke, Ulrike Lucke
„Und hier geht es übrigens jetzt gerade manuell weiter!“ 1399

Julia Niemann-Lenz, Moritz Kreinsen, Sören-Kristian Berger
Die Rolle der Informatik in der inter- und transdisziplinären Data Literacy Education 1411

Dirk von Suchodoletz, Simon Rettberg, Manuel Bentele, Steffen Ritter, Christian Rößler
bwLehrpool als offene und flexible Betriebsplattform für PC-Pools mit optionalem Fernzugriff 1417

Nicole Draxler-Weber, Henning Brink, Sven Packmohr
Barriers to Digital Higher Education Teaching during the COVID-19 Pandemic from Teachers' and Students' Perspectives 1427

Künstliche Intelligenz in der Umweltinformatik (KIU-2022)

Andreas Abecker, Julian Bruns, Stefan Naumann
3. Workshop Künstliche Intelligenz in der Umweltinformatik 1443

Günter Müller-Czygan, Jürgen Wiese, Viktoriya Tarasyuk, Ralf Tschepetzki
Energieeinsparung auf Kläranlagen mit fallbasierter Steuerung (Cased Based Reasoning) 1447

Markus Junker, Ansgar Bernardi
Entscheidungsunterstützung bei der Nutzung von Kläranlagen für Agrarsysteme 1473

Christiane Plociennik, Monireh Pourjafarian, Shehab Saleh, Tabea Hagedorn, Alice do Carmo Precci Lopes, Malte Vogelgesang, Julian Baehr, Bernd Kellerer, Maike Jansen, Holger Berg, Martin Ruskowski, Liselotte Schebek, Andreas Ciroth
Requirements for a Digital Product Passport to Boost the Circular Economy 1485

Alexander Münzberg, Christian Troost, Ansgar Bernardi
Using Transfer Learning for Quality Improved Forecasting of Temporal Agricultural Processes by Adapting Convolutional Neural Networks . . . 1495

Nina Dorffer, Julian Bruns, Andreas Abecker, Stefan Lossow
Comparison Spatio-Temporal Prediction Approaches of Point-Referenced Environmental Data 1505

Katharina Emde, Matthias Budde, Thilo Fischer, Tobias Martin, Desiree Hilbring
Interaktive Steuerung der Ausführung von KI-Algorithmen in Umweltinformationssystemen über OGC SensorThings 1517

7th GI/ACM I4.0 Workshop on Industrial Automation and Control Systems

Mahmoud Khalaf, Ludger Peters, Karl Waedt
Modeling Security Controls and System Assets As Autonomous Planning Tasks 1531

Ludger Peters, Mahmoud Khalaf, Karl Waedt, Josef Schindler, Siwar Belaidi
Model-based Integrity Monitoring of Industrial Automation And Control Systems 1539

Josef Schindler, Siwar Belaidi, Erkin Kirdan, Karl Waedt
Securing javascript runtime of OPC UA deployments 1551

Edin Kreho, Roger Djeukoua, Timothée Guiraud, Karl Waedt
Scalable backend representation of security posture of IIoT systems 1557

Jan Sudeikat, Michael Kohler-Bußmeier
Towards Integrating Multi-Agent Organizations in OPC UA for Developing Adaptive Cyber-Physical Systems 1565

Louis Roger Tchuegoue Djeukoua, Edin Kreho, Siwar Belaidi, Karl Waedt
Interactive graphical modeling of security artefacts for abstracted Industry 4.0 automation systems 1571

Yun Guo, Junjie Wang
New cybersecurity standards for IACS of the nuclear power industry in China 1583

IT-Governance und Strategisches Informationsmanagement (ITG-SIM)

Can Adam Albayrak, Andreas Gadatsch, Benedikt Haag
Agile Methoden für das IT-Controlling 1593

Sebastian Floerecke, Franz Lehner
Beschreibung von IT-Services für Kunden: Vergleich und Konsolidierung der Attribute auf Basis einer Metastudie 1613

Gian-Luca Wagner, Helmut Beckmann
Literatur-Review: State of the Art des Wertbeitrags der IT zum Unternehmenserfolg 1629

Remzi Kültür, Helmut Beckmann
Überblick über die Ansätze zur Geschäftsprozessautomatisierung 1641

14. Workshop KI-basiertes Management und Optimierung komplexer Systeme / MOC 2022

Armin Wolf
The Proportional Constraint and Its Pruning – Continued 1655

Henrik Kortum, Simon Hagen, Janis Hühn, Oliver Thomas
Analyzing Smart Services from a (Data-) Ecosystem Perspective: Utilizing Network Theory for a graph-based Software Tool in the Domain Smart Living 1667

Ulrich John
Intelligente Hochschule – Digitalisierung und Künstliche Intelligenz im Hochschulkontext 1681

Ulrich John
Intelligentes Unternehmen, Intelligenter Staat, Intelligentes Europa – Was ist zu tun? 1695

Autorinnen- und Autorenverzeichnis

Workshops

International Workshop On Digital Forensics
(IWDF)

2nd International Workshop on Digital Forensics - an Interexchange of Law Enforcement and Science

Roman Povalej[1]

The digital world is ubiquitous. The influence of digitization is constantly increasing, both in the private and in the professional world. For example, the smartphone has become our constant companion. We can network digitally, exchange ideas, access different digital sources or information systems at the same time and much more anywhere and anytime. Of course, this also has an impact on police work. A crime scene today consists of a mixture of different physical and virtual components and aspects, such as locations, people, evidence, traces, all of which are in flux and changeable over time, but also their virtual counterparts. The internet of things, smart infrastructures, smart homes, smart cities, etc. are increasingly causing investigations, traces, forensic evidence to no longer be purely physically restricted. Evidence, traces, digital traces can be copied umpteen times, transmitted or distributed multiple times and thus appear anywhere in the world where nobody has looked before.

This workshop has the goal of firstly fostering novel, creative, and applied ideas on how to extract data from information sources, analyze that data, and apply the outcome into a criminal investigation. A second goal is to form a community, interested in appalling their scientific results of computer science, data analytics, and machine learning to cyber forensics. Thus we are very pleased to offer our workshop the second time at the annual conference INFORMATIK 2022. Thanks to the contributions submitted, we were once again able to set an exciting program (in alphabetical order).

A frequentist estimation of duplicate probability as a baseline for person identification from image and video material using anthropometric measurements
This work is concerned with approaches for duplicate probability estimations derived from anthropometric measures.

A pipeline for analysing image and video material in a forensic context with intelligent systems
This work presents a realistic view on practices in digital forensics, to introduce typical image and video analysis methods used in tackling real life forensic challenges.

Challenges of Network Traffic Classification Using Deep Learning in Virtual Networks
In this paper, the authors analyse the classification rate of a Convolutional Neural Network faced with encapsulated packets.

[1] Polizeiakademie Niedersachsen., Studiengebiet 1, Aufgabenfeld Cybercrime,
Gimter Str. 10, 34346 Hann. Münden, roman.povalej@polizei.niedersachsen.de

doi:10.18420/inf2022_01

Collecting Identifying Data for Re-Identification of Mobile Devices carried at a Crime Scene using Wi-Fi Routers
As devices carried by criminals are not willing to share identifiers with a victim's home network, the authors combine collection processes for device-identifying data with exploiting viable wireless communication protocols.

Ontology in the Digital Forensics Domain – A Scoping Review
In this contribution, the state of the art of ontological formalisation in the domain of digital forensics is analysed via a scoping review.

Potential analysis for the detection of attacks on wireless networks using the Wireless Intrusion Detection System N z y m e
This paper describes the evaluation of the possibilities offered by the open-source Wireless Intrusion Detection System Nzyme.

Recommendation of Query Terms for Colloquial Texts in Forensic Text Analysis
The aim is to develop a method that recommends keywords and searches phrases based on the underlying data.

Revisiting the privacy of censored credentials
In this paper, the authors propose a system that recovers information from redacted text in raster graphics using a composition of a Convolutional Neural Network, a Recurrent Neural Network using Long short-term memory and a Connectionist Temporal Classification layer to output the most probable character sequence.

SCALA-Speech: An Interactive System for Finding and Analyzing Speech Content in Audio Data
In this work, a system for automatically finding and analyzing speech content in audio data is presented. The system provides different speech processing algorithms as well as a graphical interface (SCALA) for assisting investigators in audio analysis.

Maybe we will see you at the third edition of our Workshop at the next annual conference of the Gesellschaft für Informatik e. V. (GI) - we would be very happy to receive your contribution.

All the best and stay healthy – your committee!

Roman Povalej, Polizeiakademie Niedersachsen
Johannes Fähndrich, Hochschule für Polizei Baden-Württemberg
Dirk Labudde, Hochschule Mittweida
Heiko Rittelmeier, Sachverständiger und Trainer für Digitale Forensik
Silvio Berner, Hochschule der Sächsischen Polizei (FH)
Wilfried Honekamp, Hochschule Stralsund

Recommendation of Query Terms for Colloquial Texts in Forensic Text Analysis

Jenny Felser[1] Jian Xi[2] Christoph Demus[3] Dirk Labudde[4] Michael Spranger[5]

Abstract: Textual social media content and short messages have gained in importance as evidence in criminal investigations. Yet, the large number of textual data poses a great challenge for investigators. Even though text retrieval systems can assist in finding evidential messages, the success of the search still depends on entering appropriate search terms. However, for colloquial texts these are difficult to determine because one cannot be sure about what terms are used in the texts and might be of interest. Therefore, the aim is to develop a method that recommends keywords and searches phrases based on the underlying data. A particular challenge here is that the appropriate search terms are often non-obvious words that are not expected to be found in the data, but are particularly relevant. In total, four methods were evaluated for extracting and suggesting the most relevant terms and phrases using a real-life dataset. The best results were obtained with topic modeling considering syntagmatic relations.

Keywords: query terms; colloquial texts; forensic text analysis; text retrieval

1 Introduction

In contemporary digital age, the communication via messenger services such as WhatsApp increases significantly. Consequently, the analysis of chat messages of mobile devices also gains in importance for forensic purposes. However, the increasing number of messages develops to a great challenge for the investigators. Although text retrieval systems can assist in finding evidential information in the numerous messages, the success of the search largely depends on entering suitable query terms. Yet, these are difficult to find, especially in the case of colloquial texts. Mostly, it is not known which terms are relevant for the specific case. Furthermore, one cannot be sure if particular words are even used in the chat messages. Therefore, the aim of this paper is to develop a method for recommending search words and phrases based on the underlying dataset. For this purpose, different methods are analysed

[1] University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany jfelser@hs-mittweida.de
[2] University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany xi@hs-mittweida.de
[3] University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany cdemus@hs-mittweida.de
[4] University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany labudde@hs-mittweida.de
[5] University of Applied Sciences Mittweida, Technikumplatz 17, 09648 Mittweida, Germany spranger@hs-mittweida.de

doi:10.18420/inf2022_02

and compared, which have in common that they do not require any prior knowledge of the data.

Thereby, a particular challenge is that the most relevant terms for an investigation are often non-obvious and unfamiliar words. For instance, if the case deals with the financial support of a terrorist group, the investigator will assume that the dataset contains words such as "money", "transfer" or "account". However, it would be much more interesting if the chats mentioned names of terrorist organisations, brands of weapons that have to be bought for a terrorist attack, or places connected to the terrorist group.

One application of the proposed approaches consists in the development of a Term Tree Recommender (TTR) for the Mobile Network Analyzer (MoNA), a forensic tool for analysing chat messages [1], where the so-called Term Tree identifies messages containing relevant terms for a certain case. The purpose of the TTR is to suggest words that can subsequently be imported to the Term Tree. A time-limited trial version of MoNA including the implemented TTR is available for download [6]. In addition, the presented methods can also be applied in general to find important terms and phrases in texts and particularly in short messages.

The paper is organized as follows: Related work is presented in Sect. 2. After it, Sect. 3 shows the details of the proposed approaches. Afterwards, the experimental results are presented in Sect. 4 and discussed in Sect. 5. Finally, Sect. 6 concludes the paper.

2 Related Work

There have been only few studies to address the issue of recommending search terms in forensic texts. Besides, those studies have focused on the analysis of e-mails and have not considered short messages and chats. For instance, Koven et al. [2] presented a search term recommender for InVEST, a forensic tool for finding evidence in large e-mail datasets, where the proposed approach assumes that the investigator has already searched for a word in the data. Subsequently, the recommender suggests the most important terms in the result set using a variant of TF-IDF algorithm. Then, the investigator can inspect the suggested terms in order to discover more relevant e-mails. The proposed methodology is not applicable to our recommender system, since the search procedure within is fully conducted from scratch with its assistance, whereas the recommender developed by Koven et al. [2] is part of an iterative search process. Furthermore, Teng [3] pointed out that TF-IDF tends to extract words with high frequency and is thus inappropriate to suggest specific and uncommon words.

Addressing the second problem, Joshi and Motwani [4] introduced the recommender TermsNet with the purpose to suggest non-obvious, but nevertheless relevant terms in the field of sponsored search. The authors proposed a graph-based technique, which explores semantic associations between words by considering the syntagmatic relations between

[6] https://www.hs-mittweida.de/spranger/

concepts. However, their approach requires the user to specify a few relevant seed words, whereas in this paper, the focus is on the case that no relevant terms of the data are known.

Moreover, a large number of existing studies have examined methods for suggesting query terms in the context of web search. However, most of those approaches rely on query logs [5][6][7] and cannot be used for recommending important words based on a certain underlying dataset. A more relevant work in this field was conducted by Kubek and Unger [8], where the recommender system DocAnalyser suggests the most relevant words of a previously visited web page by a graph-based method that is based on a modified HITS-algorithm. Subsequently, the user can enter those words as query to find similar web documents. Yet, Kubek and Unger [8] emphasized that their approach achieves particularly good results, when the underlying dataset deals with one homogeneous topic. Otherwise, the recommender will mainly suggest words that are related to the topic with the highest coverage in the data [9]. This is particularly problematic for our recommender system because the case-related topic only represents a small portion of the conversations. Furthermore, the proposed method addresses term recommendations merely based on a single web page. Therefore, it is not necessarily appropriate for handling large data.

To overcome these limitations, instead of using HITS-algorithm, a further graph-based algorithm, namely "Rapid automatic keyword extraction" (RAKE) [10], is reported by Wang et al. [9] for suggesting the most important phrases of online news articles. Even though this approach yields promising results, it cannot be guaranteed that it also performs well on messages. In contrast to news articles, short messages and chats contain numerous typos, grammatical mistakes, acronyms [11] and colloquial expressions. Therefore, it constitutes a particular challenge for traditional algorithms and methods.

Keerthana [12] presented a recommender that uses topic modeling for suggesting queries based on conversation fragments. The basic idea is to recommend the terms with the highest probability in the most relevant topic of the dataset. As topic modeling has proven to be successful in revealing latent semantics [13], this method has also potential to recommend specific and unexpected words. However, previous studies [14][15] have shown that conventional topic models do not perform well on short text collections due to their high sparsity. To avoid this problem, we make use of a variant of topic modeling incorporating syntagmatic relations, as proposed by [16].

Consequently, previous studies have either failed to recommend particularly specific terms or have been constrained to the case that the user already knows some relevant words. To fill this gap, the objective of this paper is to develop a new method for term recommendation, which, in addition, is also able to cope with the special linguistic and grammatical characteristics of forensic short messages.

3 Methods

To find the most appropriate approach for recommending search terms, we compared the following four algorithms: differential analysis, document clustering, standard topic modeling and a variant of topic modeling considering syntagmatic relations. All methods were evaluated using a dataset of a real case about the financial support of a terrorist group, which includes approximately 180.000 messages of the messenger service WhatsApp stored on the mobile phone of one case-related person. In order to assess the quality of the different approaches, the recommended terms were compared with the words of an existing Term Tree that had previously been created by an investigator for the analysis of the corresponding case using the software MoNA. The Term Tree contained the following 25 words that the investigator had deemed relevant to the case: *Geld (money), Schwester (sister), Nalan, Hawala, Nierderlande (Netherlands), Buch (book), Bücher (books), scan, Seiten (pages), Kapitel (chapter), Wörter (words), Bieber (beaver), Layth, Verleger (publisher), money, Druck (print), angekommen (arrived), Dolar (dollar), €, Euro (euro), $, Western (western), Union (union), PayPal* and *Dollar (dollar)*. A method was considered as suitable if it recommended many of these or similar terms such as synonyms or thematically related words. The procedure from the preprocessing of the dataset to the evaluation of the applied approaches is shown in Fig. 1.

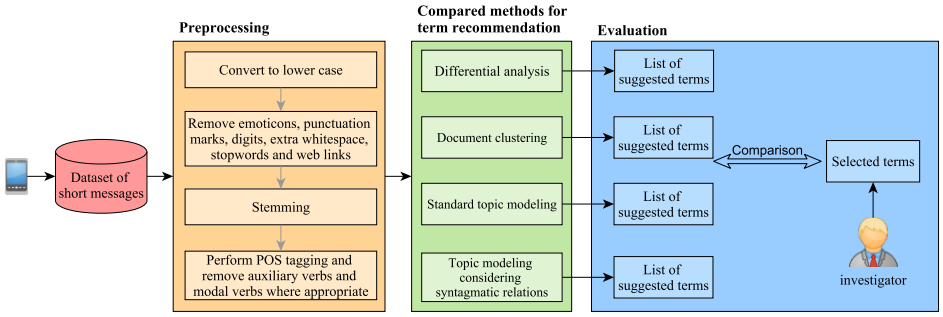


Fig. 1: Procedure applied for the comparison of methods for the recommendation of query terms.

3.1 Preprocessing

A number of preprocessing steps were performed before the comparison of the mentioned algorithms. Firstly, the messages were tokenised in terms with lower case. Emoticons, punctuation marks, digits and extra whitespace of the messages were removed. In addition, German and Turkish stopwords were removed using the stopword lists provided by Diaz [17] and all web links were also removed. Stemming was used for all methods in order to reduce the words to their root or stem forms. At recommendation step, the matched words, which are stemmed, were completed in order to recommend the original, correct terms.

Regarding topic modeling considering syntagmatic relations, all parts of speech (POS) were used. Adjectives were additionally included for differential analysis, clustering and standard topic modeling. TreeTagger was used for POS tagging [18]. Moreover, auxiliary verbs and modal verbs were removed for clustering and both variants of topic modeling. Since the dataset consists mainly of German messages, only the German TreeTagger model was used for POS tagging and the removal of auxiliary and modal verbs was also limited to German words.

3.2 Differential analysis

A first idea was to apply differential analysis [19], which is an approach of statistical text analysis [8]. Differential analysis compares the frequencies of the words in the messages with their frequencies in the so-called reference corpus, a text corpus in common German language. In this work, the corpus "Deutscher Wortschatz" [20] of the University of Leipzig was chosen as reference corpus, which predominantly consists of online newspapers. Differential analysis identifies all words that appear in the chat messages significantly more often than in the reference corpus [19]. As these words are considered as discriminating terms of the messages, they are also appropriate search term recommendations. The extracted words were sorted based on the difference of their frequencies to the reference corpus.

3.3 Document clustering

A further approach was to cluster the messages using an agglomerative hierarchical cluster analysis as explained by Murtagh and Contreras [21]. However, the high sparsity of short messages is a well-known problem with hierarchical clustering [22]. One way to tackle this problem is to form pseudo-documents by aggregating several messages [22], whereby in this work, a pseudo-document consisted of all messages written in one chat at one day. Initially, four clusters were created, of which the investigator must select the case-related one. Subsequently, only the messages in the respective cluster were considered for extracting relevant words, whereas irrelevant small talk conversations were excluded in order to reduce noise in the data. The words that occured most frequently in the considered messages were recommended, which is similar to the approach described by Jones, Doane, and Jones [23].

3.4 Standard topic modeling

Another method for explorative data analysis is topic modeling, for which Latent Dirichlet Allocation (LDA) described by Blei, Ng, and Jordan [24] was chosen in this study. The input of the LDA is usually a document-term matrix (DTM) [25]. As with the previous clustering, a document was defined as the amount of all messages that were written in

one chat at one day. The number of topics was set at four in accordance with clustering. The output of the LDA needed to recommend terms is the probability distribution over words that characterise each topic [24]. Similar to document clustering, the investigator is presented with all the topics from which he has to select the case-related one. Afterwards, the most probable words of that topic are suggested.

3.5 Topic modeling considering syntagmatic relations

As mentioned before, standard LDA can frequently be improved for short messages by taking syntagmatic relations into account [16]. To do so, the LDA was repeated with a term co-occurrence matrix (TCM) as input, which captures the local context of words by using a skip-gram model [23]. In contrast to the previously described approach, the most appropriate number of topics was estimated based on probabilistic coherence as proposed by Jones, Doane, and Jones [23]. Analogous to standard topic modeling, recommendations of query terms are obtained by suggesting the most probable terms of the topic selected by an investigator.

4 Results

4.1 Differential analysis

In general, differential analysis did not lead to the recommendation of relevant terms. As an example, the first nine terms recommended by difference analysis and their frequencies in the messages are shown in Tab. 1. As can be seen, none of the words of the Term Tree were suggested, but mostly Arabic phrases and colloquial terms.

term	term frequency
Allah	1871
Alhamdulillah	939
Shaa	872
Inshallah	844
Subhanallah	792
haha	748
hahaha	666
nochmal	621
amin	547

Tab. 1: The first nine terms suggested by differential analysis.

4.2 Document clustering

It was not possible to identify the case-related cluster because none of the most frequent terms in the respective documents also appeared in the Term Tree. Furthermore, it is remarkable that the clusters vary widely regarding the number of assigned documents. As an example, the five most frequent words and the number of documents respectively pseudo-documents in the four clusters are given in Tab. 2.

cluster	# documents	most frequent terms
1	3831	ok, diggi, digger, moin, hayirli
2	2	yorum, resimlere, çekiyor, yapmis, deyince
3	2	einzelne (individual), Preise (prices), verschiedenen (different), digga, schau (look)
4	1	gucks (look), Bunker (bunker), digga, ersma (first), frohes (happy)

Tab. 2: The five most frequent words of four clusters using hierarchical document clustering.

4.3 Standard topic modeling

The first of four topics calculated by LDA could be identified as the case-related topic, because it is the only one for which terms in the Term Tree appear among the 80 most probable words, namely “Druck” and “PayPal”. Moreover, there are seven further terms among these 80 words that are thematically similar to the terms in the Term Tree, for example “Sponsoren” and “fees”. Those words are shown with their topic probabilities and their position among the 80 most probable words in Tab. 3.

position	term	probability in topic 1
7	spenden (donate)	0.0016043884
11	Rechnung (bill)	0.0014620923
23	Druck (print)	0.0012842222
43	fees	0.0011419261
54	PayPal	0.0010707781
58	Preise	0.0010352041
64	Belege (receipts)	0.0009996300
71	Gebühren (fees)	0.0009640560
75	Sponsoren (sponsors)	0.0009284820

Tab. 3: Relevant topic calculated by traditional topic modeling.

4.4 Topic modeling considering syntagmatic relations

Estimating the appropriate number of topics by probabilistic coherence resulted in 13 topics. The ninth topic was considered the relevant topic for the case as shown in Tab. 4 because the word “Geld” of the Term Tree and further three thematically related terms, namely “Fees”, “Konto” and “Rechnung”, are among the 50 most probable words of this topic. In addition, we also regarded words such as “Twitter”, “Event” and ”Tipeee” as important for the case because it was apparent from the context that these words were related to the financial support of the terrorist organisation. As outlined in Tab. 4, a total of 15 terms were considered as relevant. Those terms are shown with their position among the 50 most probable terms and their topic probability.

position	term	probability in topic 9
1	Twitter (twitter)	0.017992353
2	Team (team)	0.015728373
4	Event (event)	0.014847936
5	Statement (statement)	0.014093276
8	Tipeee	0.012835509
9	Twitch	0.011955073
15	Discord	0.009439539
18	SWH	0.008433326
19	Fees	0.008181772
24	Partner	0.007678666
25	Geld (money)	0.006546676
26	Konto (account)	0.006420899
39	Betterplace	0.004534249
40	Projekt (project)	0.004408472
45	Subs	0.003779589
49	Rechnung (bill)	0.003402259

Tab. 4: Relevant topic calculated by topic modeling incorporating syntagmatic relations.

4.5 Comparison with state-of-the-art approaches

The results of the proposed approaches were compared with two commonly used methods for the recommendation of query terms, namely TF-IDF and RAKE. The first ten terms suggested by these methods are listed in Tab. 5. As RAKE is only applicable for finding relevant words in single documents and not in the whole dataset, the presented term recommendations refer to each pseudo-document consisting of all messages in one chat. As an example, the recommended terms for the first three chats are presented. As outlined in Tab. 5, none of the terms of Term Tree or similar words are among the words suggested by these methods.

position	suggested query terms by TF-IDF	suggested query terms by RAKE		
		chat 1	**chat 2**	**chat 3**
1	voll	ewig draußen rum	fufu danke fürs	guten appetit danke
2	enişte	absoluter ruhemodus	btw	biste glitch theater
3	güzel	guten hunger	relativ	cool lässt
4	klingt	halle geworfen	sascha	haha geilo
5	digga	leute direkt	wach	iwie ggn
6	nen	liebe messen		paar fotos
7	ding	nähe entspannt		paar vorträge
8	wach	voll happy		schmecken naja
9	hause	nacht eistüte		würd sagen
10		draußen		

Tab. 5: The first ten terms suggested by two common methods for recommending query terms.

5 Discussion

Both standard topic modeling and topic modeling considering syntagmatic relations suggested terms that were indeed considered relevant by an investigator. In contrast, other current approaches such as TF-IDF or RAKE were not able to identify the words that were actually important, but mainly recommended slang terms and empty phrases. This result highlights the suitability of the proposed approaches for chat messages.

However, standard topic modeling fails to recommend specific and non-obvious words. For instance, it is not surprising that the terms “fees” and “Gebühren”, as shown in Tab. 3, occur in the context of the topic “money”. In contrast, by incorporating syntagmatic relations, topic modeling is able to recommend specific and unfamiliar words such as the crowdfunding platform “Tipeee”, the association “Streaming with Heart” (SWH) and the donation platform “Betterplace”. Although these words have not been previously added by the investigator to the Term Tree, it could be concluded from their context in the chat messages that those words are indeed case-related. Consequently, this approach succeeds in revealing terms that the investigator would not otherwise have considered as search terms. Therefore, this method is also considered the most useful approach for recommending search terms.

Moreover, the results demonstrate that differential analysis is unsuitable for suggesting appropriate query terms based on the chat messages. A possible explanation for this result is that differential analysis does not only reveal differences of content between the messages and the reference corpus, but also stylistic differences [26]. Whereas the reference corpus is written in German standard language, the messages in this study are colloquial and contain expressions of youth language and Turkish phrases. As a consequence, the words that exclusively occur in the messages, thus being identified as characteristic terms, are particularly slang terms, anglicisms and Turkish filler words. In addition, these results

demonstrate that term frequency is not sufficient as sole criterion for extracting relevant words, since the high-frequency terms, which can be seen in Tab. 1, are not necessarily representative for the analysed text.

With regard to clustering, the inadequate results can be explained by the fact that the outcomes of clustering, especially of hierarchical clustering algorithms, strongly depend on the underlying data [27]. Algorithms for document clustering are based on documents that are linguistically correct [28]. The chat messages differ significantly of such documents.

The main limitation of our study so far is the lack of a more thorough evaluation of the compared approaches. Therefore, a more comprehensive user study is required, in which the relevance of terms to the messages is assessed by multiple annotators rather than just a single investigator. In addition, the methods should be evaluated on more forensic datasets covering different offense domains and multiple languages. Besides, it would be interesting to investigate if topic modeling considering syntagmatic relations also yields promising results for other forensically relevant document types, especially for longer text types such as case records and interrogation transcripts.

Furthermore, the approaches that have been investigated so far are completely unsupervised and assume that no important terms of the data are known. However, the investigator might already know a few relevant words, for example from police interrogations or from the case record. Hence, further studies should examine methods considering those words for recommending query terms. Possibilities include to analyze paradigmatic relations to case-related terms and to apply semantic similarity measures such as Latent Semantic Analysis and Pointwise Mutual Information.

6 Conclusion

Finding evidential information in numerous chat messages that need to be analysed in the investigative process is a challenging task. Therefore, in this paper we presented different approaches to develop a method for recommending query terms for colloquial texts. Overall, topic modeling incorporating syntagmatic relations was the method of the four investigated approaches that was capable of recommending the most terms that had previously been considered by an investigator relevant. A particular advantage of this method is that it is able to suggest non-obvious and specific terms that the investigator would not otherwise have thought to search for in the dataset. Entering these appropriate query terms in text retrieval systems prevents important clues contained in the messages from being overlooked. Furthermore, the use of specific search terms, as opposed to general terms, can also reduce the number of messages returned by those systems, resulting in significant time savings for the investigator.

Nevertheless, further experiments involving more real-life datasets and evaluation of the results by more annotators need to be conducted. Furthermore, a possible improvement is to apply methods that take into account a few known relevant words.

References

[1] Michael Spranger et al. "MoNA: Automated Identification of Evidence in Forensic Short Messages". In: *International Journal On Advances in Security* 9.1&2 (2016), pp. 14–23.

[2] Jay Koven et al. "InVEST: Intelligent Visual Email Search and Triage". In: *Digital Investigation* 18.16 (Apr. 2016), pp. 138–148.

[3] Miao Teng. "Using the Ship-Gram Model for Japanese Keyword Extraction Based on News Reports". In: *Complexity* 21.4 (2021), pp. 1–9.

[4] Amruta Joshi and Rajeev Motwani. "Keyword Generation for Search Engine Advertising". In: *IEEE International Conference on Data Mining-Workshops*. Vol. 6. IEEE. 2006, pp. 490–496.

[5] Ricardo Baeza-Yates, Carlos Hurtado, and Marcelo Mendoza. "Query Recommendation using Query Logs in Search Engines". In: *International conference on extending database technology*. Springer. 2004, pp. 588–596.

[6] Qi He et al. "Web Query Recommendation via Sequential Query Prediction". In: *IEEE International Conference on Data Engineering*. Vol. 25. IEEE. 2009, pp. 1443–1454.

[7] Qiaozhu Mei, Dengyong Zhou, and Kenneth Church. "Query Suggestion Using Hitting Time". In: *Proceedings of the ACM conference on Information and knowledge management*. Vol. 8. 17. 2008, pp. 469–478.

[8] Mario Kubek and Herwig Unger. "Search Word Extraction Using Extended PageRank Calculations". In: *Autonomous Systems: Developments and Trends*. Springer, 2012, pp. 325–337.

[9] Zihuan Wang et al. "A news-topic recommender system based on keywords extraction". In: *Multimedia Tools and Applications* 77.4 (2018), pp. 4339–4353.

[10] Stuart Rose et al. "Text Mining: Applications and Theory". In: Citeseer, 2010. Chap. Automatic Keyword Extraction from Individual Documents, pp. 1–20.

[11] Ville H. Tuulos and Henry Tirri. "Combining Topic Models and Social Networks for Chat Data Mining". In: *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI'04)*. IEEE. 2004, pp. 206–213.

[12] S. Keerthana. "Recommended Search of Documents from Conversation with Relevant Keywords Using TextSimilarity". In: *Journal of Network Communications and Emerging Technologies (JNCET)* 7.2 (Feb. 2017). ISSN: 2395-5317.

[13] Yuzhong Chen et al. “Popular topic detection in Chinese micro-blog based on the modified LDA model”. In: *Web Information System and Application Conference (WISA)*. Vol. 12. IEEE. 2015, pp. 37–42.

[14] Ou Jin et al. “Transferring Topical Knowledge from Auxiliary Long Texts for Short Text Clustering”. In: *Proceedings of the ACM international conference on Information and knowledge management*. Vol. 20. 11. Oct. 2011, pp. 775–784.

[15] Yuheng Hu et al. “What Were the Tweets About? Topical Associations between Public Events and Twitter Feeds”. In: *Proceedings of the International AAAI Conference on Web and Social Media*. Vol. 6. 1. 2012, pp. 154–161.

[16] Xiaohui Yan et al. “A Biterm Topic Model for Short Texts”. In: *Proceedings of the 22nd international conference on World Wide Web*. Vol. 13. 22. 2013, pp. 1445–1456.

[17] Gene Diaz. *Stopwords ISO*. `https://github.com/stopwords-iso/stopwords-iso`. Sept. 2020.

[18] Helmut Schmid. *TreeTagger-a part-of-speech tagger for many languages*. Tech. rep. Ludwig-Maximilians-Universität Munich, 1994.

[19] Gerhard Heyer, Uwe Quasthoff, and Thomas Wittig. *Text Mining: Wissensrohstoff Text - Konzepte, Algorithmen, Ergebnisse*. Bochum: W3L-Verlag, 2006. ISBN: 978-3-937-13730-8.

[20] Dirk Goldhahn, Thomas Eckart, and Uwe Quasthoff. “Building Large Monolingual Dictionaries at the Leipzig Corpora Collection: From 100 to 200 Languages”. In: *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*. Vol. 12. 8. 2012, pp. 759–765.

[21] Fionn Murtagh and Pedro Contreras. “Algorithms for hierarchical clustering: an overview”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 2.1 (Jan. 2012), pp. 86–97.

[22] Oren Tsur, Adi Littman, and Ari Rappoport. “Efficient Clustering of Short Messages into General Domains”. In: *Proceedings of the International AAAI Conference on Weblogs and Social Media*. Vol. 7. 2013, pp. 621–630.

[23] Thomas Jones, William Doane, and Thomas Jones. *Package textmineR*. `https://cran.microsoft.com/snapshot/2022-04-10/web/packages/textmineR/textmineR.pdf`. June 2021.

[24] David M. Blei, Andrew Y. Ng, and Michael I. Jordan. “Latent Dirichlet Allocation”. In: *Journal of Machine Learning Research* 3.1 (Jan. 2003), pp. 993–1022.

[25] Kurt Hornik and Bettina Grün. “topicmodels: An R package for fitting topic models”. In: *Journal of statistical software* 40.13 (2011), pp. 1–30.

[26] Uwe Quasthoff. “Korpusbasierte Wörterbucharbeit mit den Daten des Projekts Deutscher Wortschatz”. In: *Linguistik online* 39.3 (June 2009), pp. 151–162.

[27] Moses Charikar, Vaggos Chatziafratis, and Rad Niazadeh. “Hierarchical Clustering better than Average-linkage”. In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2019, pp. 2291–2304.

[28] Kevin Dela Rosa et al. “Topical Clustering of Tweets”. In: *Proceedings of the ACM SIGIR: SWSM* 63.10 (2011).

Potential analysis for the detection of attacks on wireless networks using the Wireless Intrusion Detection System Nzyme

Eisenhut, Maximilian[1], Honekamp, Wilfried[2]

Abstract: Due to the flexibility and low cost of acquisition compared to wired network connections, wireless networks continue to proliferate. Due to this increasing number and the characteristics of a shared medium, it offers potential attackers a suitable platform to easily gain access to diverse network types. To this end, the range of specialised hardware and software for attacking wireless networks is constantly evolving. Information on the location and other parameters of wireless networks is also documented and updated online in a largely automated manner. Particularly in the economic as well as in the public environment, a special need can thus arise to detect attacks, identify attackers and initiate countermeasures on the basis of this information. This paper describes the evaluation of the possibilities offered by the open-source Wireless Intrusion Detection System (WIDS) Nzyme. For this purpose, the messages that occur during different attacks were examined. Furthermore, real data was recorded and evaluated based on the parameters from the test attacks to draw conclusions about the type and frequency of attacks. The ratio between legitimate reports and false alarms was also determined. Test attacks were successfully detected and could be assigned to possible attacks. Real data was recorded at three locations and compared with the patterns from the test attacks. The evaluation shows that the rate of false alarms in real operations is unacceptable, at over 27%. The causes for this are mostly misconfigurations and atmospheric disturbances. The study further shows, that combined alarm messages allow conclusions to be drawn about the type of attack carried out and thus the number of false alarms can be reduced. The effort and benefit of a WIDS are currently not yet in a meaningful relationship. Nevertheless, use and further development are recommended, taking these circumstances into account.

Keywords: Anomaly detection, intrusion detection system, Nzyme, WLAN.

1 Introduction and Background

Due to the advantages of wireless local area networks (WLAN) in terms of flexibility and cost-effectiveness compared to wired networks, they are becoming more and more widespread, whether as public access points (AP) to the Internet, for automation in the business environment, or for use by public authorities [Wi22a]. Bandwidth and speed are increasing due to the further development of technologies. The aspect of security is also

[1] Hochschule Stralsund, Fakultät für Elektrotechnik und Informatik, Zur Schwedenschanze 15, Stralsund, 18435, maximilian.eisenhut@hochschule-stralsund.de
[2] Hochschule Stralsund, IACS, Zur Schwedenschanze 15, Stralsund, 18435, wilfried.honekamp@hochschule-stralsund.de, https://orcid.org/0000-0003-2931-7047

doi:10.18420/inf2022_03

subject to constant development through the implementation of new security standards and protocols for the encrypted transmission of data.

On the other hand, there is the growing complexity of the IT infrastructure and the associated potential for vulnerabilities in these systems. In addition, there is the increasing public availability of information on the nature and geographical locations of wireless networks worldwide [Wi22b]. The increasing supply of hacking hardware and software also favours the abusive use of this technology. Two fundamental interests collide in this area: on the one hand, the goal of finding vulnerabilities in systems through observation and attacks and exploiting them for set goals. And on the other hand, to be able to take appropriate countermeasures to protect the basic goals of information security, confidentiality, integrity and availability. Thus, there is a constant interplay between attack and defence. "WiFi will continue to be more vulnerable to attack than hardwired LAN as long as electromagnetic radiation fails to obey property lines." [Be05], (p. 28)

One possible defence measure here can be the use of a system that checks network traffic for anomalies and stores evidence data. This is important for a forensic investigation during or after an incident. Especially for wireless networks, due to their physical nature, there may be a relevance to use such so-called wireless intrusion detection systems (WIDS). Different systems are offered on the market, e. g. Graylog, Nzyme, or Wirebug. The possibilities of data evaluation from intrusion detection systems (IDS), for this purpose, are described, among others, in the article by Kasongo and Sun "A Deep learning method with filtered based feature engineering for wireless intrusion detection system" [Ka19]. The use of an explicit data set, for machine learning based on data mining, is not intended to be the content of this work. Instead, this work addresses the current relevance of WIDS and discusses whether a dedicated open source WIDS can contribute to the security of wireless networks.

An IDS is a device or software that attempts to detect network attacks in order to prevent possible incidents/attacks by collecting and analysing data. To ensure security, IDS have already been developed for use in the WLAN, known as WIDS. Similar to traditional IDS, these WIDS can detect patterns of known attacks, identify anomalous network activity, and detect WLAN policy violations by monitoring and analysing network, user, and system activity. Like traditional signature-based IDS and anomaly-based IDS, WIDS can detect anomalies according to either pre-defined signatures (known threats) or observed abnormal (exceeding baselines) network behaviour. WIDS can be either centralised or decentralised. In a centralised WIDS, the central management system combines and analyses all wireless data from each distributed individual sensor. In a decentralised WIDS, there is more than one device that both collects data and generates the intrusion alerts by analysing the data. [Ya13]

2 Method

The suitability of an open source WIDS in practice is to be examined by means of a case study using Nzyme as an example. It is a decentralised open-source solution, and thus meets the requirements to be operated independently at different measurement sites. The system is designed as an explicit wireless intrusion detection system. The software runs on Raspberry Pi OS or Ubuntu Server, is implemented in Java and uses a PostgreSQL database.

The relationship between legitimate alerting and the output of false alarms will also be examined. The differentiation of the attacks is done by comparing the recorded data with alarm patterns detected during test attacks. In order to check whether a conclusion can be drawn about the executed attack based on differently occurring alarm messages or combinations of these, the attack categories spoofing, denial of service and deauthentication were recorded as manual attacks. Furthermore, the attacks were carried out with dedicated hacking hardware in order to differentiate them from automated attacks. In addition, a Kr00k attack was executed on a matching client device to further differentiate the detection. The Kr00ker attack is a variant of the "Krack" attack, which exploits a vulnerability in the 4-way handshake for authentication between the client device and the base station. Both devices use a pairwise master key (PMK). In this case, the signal is delayed by an attacker so that the ascending packet number between the client device and the server differs greatly, causing the connection to be reset. In the process, the PMK is set to 0. This allows the attacker to read the entire traffic without encryption.

To represent different applications and environments of WLAN, monitoring was done at three locations (university, local government, private residence) and the alarm messages that occurred were recorded. After installing the monitoring stations, a pilot study was conducted over two days for configuration. During this time, the network traffic occurring at the site, the legitimate fingerprints, channels and basic service set identifier (BSSID) of the AP, and the thresholds of the de-authentication frames were recorded. The average power level of the monitored AP was recorded to provide a reference value for detecting deviations, such as those caused by rogue AP. Based on this, false alarms can be minimised and it can be determined which attack messages or combinations, occur in which attack. The minimisation of false alarms plays an important role, since too frequently occurring false alarms, lead to a decrease of the sensitivity of the observing place, and so with a genuine alarming, by the observing person, no action takes place. [Gr66]

To perform the test attacks, airgeddon [V117] was run as a bash script on a PC running Kali Linux [Of22]. This is freely accessible, largely automated, and covers multiple attack patterns, such as current denial of service, de-authentication, offline decrypt, evil-Twin, and enterprise attacks. The Kr00ker attack was executed based on a proof-of-concept script. Attacks using a copied specific BSSID were carried out manually using air-base-ng and macchanger. To complement this, a fake AP was set up [Ha14].

The AP to be monitored can be stored with fingerprint, channel, BSSID and SSID. A threshold value for the number of de-authentication frames to trigger an alarm message can be defined. The address and access data of the system can be stored for alerting via e-mail. The address and access data for an existing logging system can be stored in the configuration for the central recording of data. This is used to be able to carry out general evaluations independent of location. To record the data, 3 measuring stations were set up. Two based on a Raspberry Pi 4 and one based on a Raspberry Pi 3B. A tracker for the localisation of possible rogue AP was set up on the basis of a Raspberry Pi 4 and connected via compatible long range (LoRa) modules over LoRa WAN with the measuring station ECTO-1 (Figure 1) located at the university.



Figure 1: Nzyme receiving station ECTO-1

As a wireless network adapter, a USB stick with RT5572 chipset and two antenna connections, was operated in monitor mode to monitor the defined networks in the 2.4 GHz and 5GHz bands. In monitor mode, no connection is established to a specific AP, but all data packets within range are picked up. To increase the range, the device was connected to two 9dBi antennas. The network adapter with two antennas simulates as "beacon_trap" an AP with SSIDs without underlying infrastructure to provoke attacks. If this network is offered by a device, it is an attack. Another USB device with an antenna simulates a cell phone that is to be used to connect to known networks. For this purpose, fictitious SSID are requested by the device. As soon as these are offered by a rogue AP within range of the WIDS, the "probe_response_trap" is triggered and issues a corresponding alarm message. Alarms provoked in this way can be localised via the fingerprint of the rogue AP using a tracker. The tracker is connected to the measuring station via LoRa Wan. The background for this is the low susceptibility of LoRa WAN to jamming attacks on WLAN. [Ös18]

As an example, a tracker based on a Raspberry Pi 4 with RT5572 USB WLAN adapter and LoRA WAN module, in combination with a tablet as display, was set up for testing purposes. The localisation of suspicious devices, such as rogue AP, is implemented on

the basis of a measurement of the recorded power level. The measurement is in decibel milliwatts (dBm); the lower the negative value, the stronger the signal or the shorter the distance to the transmitter [Ri21]. Two further measuring stations were implemented as ECTO-2 in a domestic environment and ECTO-3 in an official environment. These were operated without "trap" and tracker, as a pure base station, in order to further differentiate the accruing data situation. Incoming messages were recorded on a logging server using the open-source software Graylog. All incoming messages were recorded. For data protection reasons, the data on other wireless networks is not evaluated. In production operation, MAC addresses of client devices and new AP would be recorded. The recorded messages are sorted by timestamps. Timestamps without messages are discarded. Subsequently, implausible messages are removed. Implausible messages are alarm messages for which only one message occurred within half an hour or which are contradictory.

For the evaluation of the alarm messages, they are filtered in the first step and separated according to the respective measuring point. Entries without message or single entries, except with the message "deauth_flood", are removed. The sum of the individual messages by type and number is recorded. Entries with multiple simultaneous messages are grouped and summed. Entries with an interval of less than thirty minutes are grouped and summed. In order to draw conclusions about the attack carried out, the entries are compared with the recorded patterns of the test attacks. Based on this, the number of false alarms and the type, duration and frequency of authentic alarm messages can be determined. This comparison serves as a basis for making a statement about the sensible use of the WIDS Nzyme.

3 Results

When the attacks started, the data was recorded and the corresponding alarm patterns were assigned. In the time period from 1/2/2022 to 3/6/2022, a total of 79,897,248 frames were recorded using the Graylog software [Gr19]. The majority of these, 79,875,714, were beacon frames from AP within range of the WIDS. They do not contain any anomalies.

Table 1: Recorded spoofing patterns

	beacon_rate_anomaly	crypto_change	multiple_signal_tracks	unexpected_bssid	unexpected_channel	unexpected_fingerprint

Spoofing: other encryption, other channel, same SSID, other BSSID		x		x	x	
Spoofing SSID: different encryption, same channel, same SSID, different BSSID		x		x		
Spoofing SSID: different encryption, same channel, same SSID, same BSSID	x	x	x			x
Spoofing SSID: same encryption, same channel, same SSID, same BSSID*.	x		x			x
Client device spoofing*						

*manually executed attacks

The beacon frames contain, among other things, parameters on SSID, signal strength, channel and supported encryption at the time of recording. The number of recorded alarm messages is 21,534, i. e. 0.027% of the total number of recorded messages. The alarm message "unknown_ssid" occurred 491 times. This message occurs when an SSID is detected for the first time in the reception range of the WIDS. In the process, the beacon frames of WLAN that are not monitored by WIDS are captured, which is of no further relevance for this work leading to a remaining no. of 21,043 (approx. 31% of ECTO-1, 2% of ECTO-2, and 67% of ECTO-3) alarms that can be seen in Table 2. The use of the alarm message in conjunction with other messages to detect an attack is also not relevant, as other alarm messages such as "unexpected_fingerprint_beacon" (used fingerprint is not in the expectation list) occur. It should also be added that beacon frames allow conclusions to be drawn about personal data. Thus, the protection of personal data is another aspect to reject these alarm messages. Table 1 describes the alarm messages that occurred simultaneously depending on the attack that was carried out. The rows contain the respective attack and the columns the alarm message that occurred.

The results in Table 1 show that different spoofing attacks generate different patterns. Thus, the alarm message "crypto_change_beacon" or "crypto_change_response", in combination with other occurring alarms, can be assumed as a sure sign for a rogue AP in range. For example, a spoofing attack with different encryption, different channel, same SSID, different BSSID simultaneously generated the "beacon_rate_anomaly" (beacon frames are sent in a higher frequency than expected), "crypto_change", "unexpected_bssid" and "unexpected_channel" messages. A spoofing attack with different encryption, same channel, same SSID, different BSSID, generates the messages "crypto_change" and "unexpected_bssid". The spoofing attack with different encryption, same channel, same SSID, same BSSID, generates the messages: "beacon_rate_anomaly", "crypto_change", "multiple_signal_tracks" (more than one signal track is transmitted), "unexpec-ted_fingerprint". The spoofing attack with same encryption, same channel, same SSID, same BSSID generates the messages: "beacon_rate_anomaly", "mul-tiple_signal_tracks", and "unexpected_fingerprint". Client

device spoofing with the same encryption and Mac address of an end device connected to the AP did not generate any message in the WIDS, but caused disconnections that can be noticed by the user. The denial-of-service attacks generated the message "deauth_flood" over the duration of the attack. In contrast, the de-authentication attack to record handshakes only recorded the "deauth_flood" message selectively in a timestamp.

For attacks with dedicated devices, such as Pineapple and Pwnagotchi, the messages "bandit_contact" were recorded in addition to the patterns described above. During the deauthentication attack by a Pwnagotchi, the message "pwnagotchi_advertisement" was additionally recorded. When the Krack and Kr00k attacks were launched, a "deauth_flood" message was recorded during the test attacks. No other messages could be recorded because no device with an appropriately vulnerable chip was available. Fifteen different individual types of alarms were recorded. As the test attacks show, the messages can be divided into secure and insecure types. Secure types are alarm messages that can occur individually and clearly signal an attack in range. An example of this in this work is "bandit_contact". Bandit describes hardware that is explicitly used for spying, jamming and manipulating.

Table 2: Number of alarm messages recorded

Alarm message	Quantity
unexpected_bssid_proberesp	7.248
unexpected_fingerprint_proberesp	3.937
unexpected_channel_proberesp	3.155
unexpected_channel_beacon	1.915
unexpected_bssid_beacon	1.616
beacon_rate_anomaly	702
unexpected_ssid_proberesp	653
crypto_change_proberesp	390
unexpected_fingerprint_beacon	384
deauth_flood	383
bandit_contact	356
probe_failure	163
crypto_change_beacon	68
unexpected_ssid_beacon	50
multiple_signal_tracks	23
Total	**21.043**

In the following, alarm messages that cannot be clearly assigned to an attacker or that only define an alarm in combination with other messages occurring at the same time are defined as unsafe. Table 2 shows the distribution of alarms recorded without further checking for possible duplicate or false alarms. In order to be able to make well-founded statements about alarms that have occurred, it is therefore necessary to further investigate and clean the data to exclude false-positive messages. The basis for this is the knowledge gained from the test attacks. Furthermore, it is not clear which alarm messages were recorded at which location and whether they can be correlated with each other for attack detection. In order to draw conclusions about possible attacks and to be able to classify the data according to relevance, it makes sense to correlate the data according to location and time of recording. It also makes sense to look at individual alarms in order to identify attacks that can be clearly classified as attacks or false alarms.

Alarm messages, considered individually, can be divided into the following two categories:

- Authentic alerts, which clearly mark an attack or can be identified as false alerts depending on the alert message when they occur individually.
- Messages that can only indicate an attack in conjunction with messages that occur at the same time (based on the findings from the test attacks).

Authentic alarms are "bandit_contact", "multiple_signal_track", "unexpected_fingerprint" and "unexpected_bssid". The messages "crypto_change", "unexpected_ssid", "unexpected_channel" and "beacon_rate_anomaly" in combination with other messages indicate an attack. The message "deauth_flood" must be considered separately, as it can indicate a deauthentication attack to record a 4-way handshake or, in conjunction with "unexpec-ted_fingerprint", an "Evil Twin".

4 Discussion and Conclusions

From the recorded data and the evaluation, it can be seen that a WIDS offers a possibility to significantly increase the security of WLAN. In summary, the messages "bandit_contact", "multip-le_signal_track", "unexpected_fingerprint" and "crypto_change", and thus four of the nine messages considered, can be regarded as a sure sign of an attack. The other messages cannot be considered safe on their own. These are the messages "deauth_flood", "unexpected_channel", "unexpec-ted_bssid", "unexpected_ssid" and "beacon_rate_anomaly". This increases the risk of false alarms. Careful configuration of the WIDS can minimise the risk, but not completely eliminate it. The configuration requires knowledge of the structure and operation of a WLAN, possible attacks on it and the detection of these attacks.

With this knowledge, the combination of individual alarms into patterns can further reduce the risk of false alarms. In addition, these combinations can be considered as a

basis for machine learning or the use of other technologies, such as a Bayesian network. Basically, the rate of false alarms of approximately 28% is unacceptable for the use of the tested system. With an immense number of false alarms, the attention span of humans is too short, loosely based on the fable "The Shepherd Boy and the Wolf". In addition, the time required to manually identify an alarm as an authentic message or as a false alarm is too long. This is not in proportion to the number of confirmed alarms, so this process must be automated. Machine learning helps to reduce the rate of false alarms and improve the quality and speed of detection of attacks. By combining this technology with Nzyme, the advantages of the high detection rate can be combined with the possibilities offered by Nzyme. In this way, a device can be reliably identified as an attacker and tracked down with a tracker. In order to be able to make a reproducible statement about detected attacks versus attacks that have actually taken place, a separate observation under laboratory conditions is necessary.

The use of AP to determine the position can be manipulated and thus people can be misled. A WIDS can help to detect these attacks. Attacks on WLAN in industrial and public areas can have an impact on the control networks in public and industrial companies. The use of a WIDS helps to detect these attacks and to avoid disturbances. The distribution of alarm messages over days of the week may indicate that attacks are carried out by persons working with the system. The questions of motivation and background, as well as frequency and distribution, are the subject of divergent work. The question also remains open whether and how successful key reinstallation attacks such as Krack or Kr00K are detected by wireless intrusion detection systems, which can be the subject of future work. The possibilities of evaluating commercial WIDS also offer potential for further work. In summary, the use of the WIDS Nzyme only has the potential to contribute to security when the number of false alarms is significantly reduced.

The findings of this work can form the basis for future work. For example, a comparison of different WIDS would be desirable, or the analysis of attacks on WLAN with WPA-Enterprise such as eduroam. Further cross-location evaluations are conceivable. The added value of WIDS for recording attacks on monitored networks clearly lies in the procurement of information. How this information is used in further processes depends on the respective area of application, but requires corresponding clarification.

Bibliography

[Be05] Berghel, H., Uecker, J.: WiFi attack vectors. In: Commun. ACM 48 (8), pp. 21-28. 2005.

[Gr66] Green, D. M., Swets, J. A.: Signal detection theory and psychophysics. John Wiley, New York, 1966.

[Gr19] Graylog, Inc: The Graylog Advantage. 2019. URL graylog.org/resources/the-graylog-advantage (accessed 29.04.2022).

[Ha14] Haralson, C.: How To: Create A Fake Access Point On Kali Linux (Rogue AP MItM Attack). 2014. URL youtube.com/watch?v=HePt2J4uSno (accessed 29.04.2022).

[Ka19] Kasongo, Sydney Mambwe; Sun, Yanxia (2019): A Deep Learning Method With Filter Based Feature Engineering for Wireless Intrusion Detection System. In: IEEE Access 7, pp. 38597-38607. 2019.

[Of22] Offensive Security: Kali. The most advanced Penetration Testing Distribution. 2022. URL www.kali.org (accessed 29.04.2022).

[Os18] Öst, A.: Evaluating LoRa and WiFi Jamming. Student thesis. Mid Sweden University. Faculty of Science, Technology and Media, Department of Information Systems and Technology. 2018.

[Ri21] Rittelmeier, H.: Ortung eines WLAN-Clients im Selbstbau: Möglichkeiten und Grenzen. In: Honekamp, W., Povalej, R., Berner, S., Fähndrich, J., Labudde, D. (Eds.): Polizei-Informatik 2021, pp. 19-27

[V117] v1s1t0r1sh3r3 (2017): airgeddon. A multi-use bash script for Linux systems to audit wireless networks. github.com/v1s1t0r1sh3r3/airgeddon (accessed 29.04.2022).

[Wi22a] wigle.net: Statistics. WiFi Networks Over Time. 2022. URL wigle.net/graph-large.html (accessed 29.04.2022)

[Wi22b] wigle.net: All the networks. Found by Everyone. 2022. URL wigle.net (accessed 29.04.2022)

[Ya13] Yang, C., Gu, G.: Security in Wireless Local Area Networks. In: Chen, L., Jiahuang, J., Zihong, Z. (Eds.): Wireless Network Security: Theories and Applications. Springer Berlin Heidelberg, pp. 39-58. 2013.

Revisiting the privacy of censored credentials

Viktor Garske[1] Andreas Noack[2]

Abstract: On the internet, you find numerous images like screenshots where secret parts are hidden with irreversible redaction techniques like pixelation or blurring. In this paper, we propose a system that recovers information from redacted text in raster graphics using a composition of a Convolutional Neural Network (CNN), a Recurrent Neural Network (RNN) using Long short-term memory (LSTM) and a Connectionist Temporal Classification (CTC) layer to output the most probable character sequence. We furthermore show that our model operates in an automated pipeline, performs on blurred images without modification and is even able to compensate JPEG quality loss. Finally, our test results indicate that a generic neural network can be trained successfully to assist the recovery of pixelized or blurred information on screenshots or high-quality photos.

Keywords: Neural networks, privacy, machine learning, computer vision, password, credentials, pixelized

1 Introduction

On different platforms on the internet, users utilize images like screenshots to explain complex issues or just to make things clear. In some cases, however, those screenshots contain credentials like usernames, addresses or even passwords that the uploaders do not want to share with everyone else. This is where redaction techniques like pixelation or blurring come into play, to keep the credentials secret.

Image manipulation tools like *GIMP* provide functions to redact information in photos or screenshots conveniently. From a security point of view, pixelation of data is oftenly less secure than uploaders think, as shown in previous work by Garske; Noack [GN22].

In this paper we analyze the redaction quality of Gaussian Blur and investigate whether it delivers similar weak results than pixelation or can keep the redacted data secret. Our working hypothesis is that Gaussian Blur performs even less secure than pixelation. Although pixelation and blurring are both lossy functions, blurring makes intense use of local redundancy which probably makes it easier to extract remaining information.

We contribute a generic anti-redaction system consisting of a preprocessor and a set of neural networks that is able to extract information from redacted image snippets. Our system

[1] University of Applied Sciences Stralsund, Department of Electrical Engineering and Computer Science, Zur Schwedenschanze 15, 18435 Stralsund, Germany, viktor.garske@hochschule-stralsund.de
[2] University of Applied Sciences Stralsund, Department of Electrical Engineering and Computer Science, Zur Schwedenschanze 15, 18435 Stralsund, Germany, andreas.noack@hochschule-stralsund.de

doi:10.18420/inf2022_04

provides the user with the most probable character sequence of the redacted data while using a Convolutional Neural Network (CNN), a Recurrent Neural Network (RNN) with Long short-term memory (LSTM) and a Connectionist Temporal Classification (CTC) layer in its backend. Through the use of a preprocessor, our system is resilient to JPEG quality loss. Additionally, this paper contributes a security comparison between pixelation and Gaussian Blur for different pixel block sizes and blurring radii.

The organization of this paper is as follows: Section 2 deals with related and previous work on the topic of recovering information from redacted images. We give an overview about the Gaussian Blur technique in Section 3. Section 4 introduces our approach to gain information from pixelized character sequences using a system of neural networks. In Section 5, we demonstrate how accurate our solution is. Finally, Section 7 concludes this paper and provides an outlook.

2 Related Work

In the literature, several methods for recovering redacted information have been discussed. The recovering algorithms can either be classified by their target, e.g. faces, license plates or text images or by their technique, e.g. brute-force or neural network approaches.

One big research field deals with deblurring of objects in images and videos. Blurring can occur inadvertently as a result of camera shaking, that is why the motivations for related work are multifaceted and not only privacy related. Xu et al. [Xu17], for example, use generative adversarial networks (GAN) with novel training losses for deblurring faces and text in images. In addition to it, they compare several alternative methods and state that they achieve better results for finer details. Menon et al. [Me20] contribute the PULSE algorithm enabling reliable reconstruction of pixelized faces in images using GANs and Latent Space Exploration. Their system enables users to output realistic images in high-resolution. Our approach differs from GAN approaches as we do not aim for an improved image but predict the underlying text sequence.

A forensic use case for the recovery of text in images arises from low-resolution images of CCTV. Kaiser et al. [Ka21] evaluated opportunities for text recovering from license plates, especially from low-quality JPEG images. They also incorporate neural networks in their work.

Machine Learning is not necessarily required for such tasks: Cavedon et al. [CFV11] studied possibilities for full reconstruction of pixelized parts in videos. This is possible because the redacted part under the pixelation mask moves in videos, hence disclosing additional information that can be used for the reconstruction with a Maximum a Posteriori approach. They tested out their solution by recovering license plates and faces, where they were able to reveal information believed to be hidden.

McPherson et al. [MSS16] investigate the influence of deep learning techniques against image obfuscation. They experiment with specific datasets like the MNIST digits and demonstrate information recovery. Other than in our work, they do not focus on sequences of characters.

Hill et al. [Hi16] take such character sequences into account whereby their model is based on Hidden Markov Models (HMM) instead of neural networks. While their model is designed for pixelized images, they propose an additional pixelation step for blurred images in order to process them, too. This process differs from our work, because our model learns directly from blurred images without another pixelation step.

3 Redaction using Blurring

Blurring is one of several methods to obfuscate text in images. As with pixelation, the goal is to reduce information up to a point where no human can recognize the redacted text without raising too much visual attention.

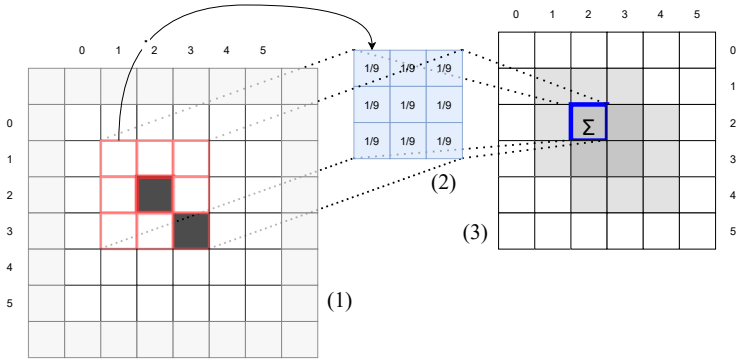


Fig. 1: Blurring

Fig. 1 illustrates the process for creating the blurred image. In order to compute the color value for a pixel on the position (x, y) in the redacted image, the algorithm averages the color values around the pixel (x, y) in the preimage.

$$\mathbf{S}_i \rightarrow \mathbf{S}_o : s_o(x, y)_m = \sum_{u=0}^{m-1} \sum_{v=0}^{m-1} s_e(x + \frac{m-1}{2} - u, y + \frac{m-1}{2} - v) \cdot h(u, v) \qquad (1)$$

The process can be summarized with e.g. Equation (1) (based on [Ni19, pp. 132 ff.]). Here, the pixel values of an input image S_i shall be mapped to an output image S_o. By selecting a proper m and kernel $\mathbf{H}$ with function $h(\cdot, \cdot)$, the output image will appear blurred. $\mathbf{H}$ and m depend on the selected technique. Fig. 1 shows a simple *box blur* where all eight pixels

around a preimage pixel are averaged evenly, so that edges appear softer and contents will be harder to read. Another very common technique is the *Gaussian Blur* with $m = 3$ and the following kernel $\mathbf{H}$.

$$\mathbf{H} = \frac{1}{16} \cdot \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \tag{2}$$

The color values of some pixels are taken more into account than others. To make an image look blurrier, the radius and kernel size are increased accordingly. As the Gaussian Blur is based on the two dimensional Gaussian distribution, the kernel matrix $\mathbf{H}$ can be calculated with Equation (3).

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{3}$$

After blurring, edges in the image appear smoother and contents are harder to read. This property is often desired to reduce image noise and details. In practice, blurring is provided by many implementations: image manipulation software like GIMP, ImageMagick or PIL offer different filters for image blurring.

4 Recovery of Redacted Information

The basic assumption for high-quality images is always the same: all redaction algorithms behave deterministically, i.e. two identical inputs always result in the same output. If we know the parameters for font family, font size, font weight or text color, we are able to compile a database to determine which input could have been created a given output. A naïve approach could incorporate a brute-force algorithm trying out all possibilities for a particular output sequence. This, however, requires huge resources. Machine Learning models can assist this process by predicting correct sequences for given blurred text images. In order to do so, the required underlying model has to be fitted to a particular training set which needs to match to the task.

Garske and Noack [GN22] create and demonstrate such an approach for pixelized images using a generic pipeline which is not tied to a specific redaction technique. This paper is based on that pipeline and neural network architecture (see Fig. 4 in Appendix). We investigate whether this approach also works for images that are not redacted by pixelation but by blurring.

The presented approach consists of two main components that aim to extract as much information as possible about the underlying text of blurred images:

1. A **preprocessor** that (a) converts pixelized images to grayscale and (b) aligns the image to left and center, and crops it to a uniform height. This is an important step to increase the accuracy of the following neural network.
2. A **neural network** including a CNN and RNN with the CTC loss function builds the core of our approach.

The proposed system generates key candidates that restrict the search space in order to optimize key recovery attacks.

4.1 Technical Remarks

The experiment is designed to approximate a real-world scenario as good as possible. With our training pipeline, text can be synthesized, blurred and saved closely to a real world sample like a screenshot that occurs with similar parameters and quality characteristics.

Furthermore, we focus on common passwords as training and test data to ensure that our neural network performs well on redacted credentials. We use a standard wordlist from the *SecLists* project by Miessler et al. [MHg21]. The selected wordlist is named `10-million-password-list-top-100000.txt` and contains the first 100,000 entries of a collection compiling the 10 million most used passwords, whereby we decided to only use the top 40,000 entries for performance reasons. The entries of the wordlist involve combinations of letters, numbers and special characters in one word.

In our redaction pipeline, text image synthesis is performed by the Python *Pillow* package [LCC21]. We use *GIMP* and the *PNG* file format with compression level 6, whereby level 1 has the best speed and level 9 the best compression rate. *PNG* is commonly used for lossless storage of high-quality images like screenshots.

4.2 Methodology

Our goal is to further analyze the previous presented pipeline and model. We want to investigate how a different redaction technique than pixelation and specific parameters like a different file format influence the performance of the prediction. In our study, blurring acts as the redaction technique and JPEG with its different quality levels as the file format.

The pipeline is able to automate the whole training process with just a set of parameters like font size or a particular wordlist. Other than some brute-force approaches, our model does not build on pixel-level analysis but works on a more abstract level due to its neural network backend. However, we want to determine whether the aforementioned neural network is able to predict the correct letters and digits with confidence when dealing with blurred images. For our experiments, we use GIMP 2.8.22 and the Gaussian Blur filter. All 40,000 training

samples, each with dimensions of 256×32 pixels, are combined together to a collage with 51200×6400 pixels. See Fig. 4 in Appendix for the exact model configuration.

First, we analyze how the *blur radius* affects our results. Therefore we create a training set for each blur radius, fit a model using our neural network architecture and evaluate it using a test set. By setting the same seed for the random function, we ensure that the collages only differ in their blur radius.

Secondly, we will conduct this experiment three times in order to examine the role of the preprocessor. In the first pass, our preprocessor works as usual. The preprocessor aligns the image using the previously mentioned algorithm and applies an additional Gaussian Blur with radius 1 to remove noise, performed by the Python PIL library. In the second pass and third pass, we disable components of the preprocessor, i.e. the images will not be aligned and/or blurred. In each case the preprocessor measures the width of the text because this is necessary for the CTC layer of our neural network model.

5 Results

Like described in subsection 4.2, we analyze every radius three times: with full preprocessing (FPP), preprocessing without aligning and blurring (NAB = no aligning and blurring) and preprocessing without blurring but with aligning (NB = no blurring). Table 1 shows the Label Error Rate (LER) and the Sequence Error Rate (SER) for different blurring radii and each preprocessor option. A blur radius m means a radius with m pixels horizontal and m pixels vertically, or $m \times m$ pixels. As additional experiment, table 2 shows the LER and SER for different JPEG quality levels when a blur radius of 10×10 pixels is used.

In order to evaluate the effectiveness, two established metrics are used: the Label Error Rate (LER) as already proposed by Graves et al. [Gr06] and the Sequence Error Rate (SER) by Soullard et al. [SRP19]. Both metrics are also used in [GN22]. The Label Error Rate (LER) metric measures the **mean normalized** edit distances between classifications and preimage labels and is defined as follows:

$$LER(h, S') = \frac{1}{|S'|} \sum_{(\mathbf{x},\mathbf{z} \in S')} \frac{ED(h(\mathbf{x}), \mathbf{z})}{|\mathbf{z}|} \tag{4}$$

whereby the LER is computed for a temporal classifier h (i.e. the model) and a given test set S' that is disjoint from the training set S. x stands for the input sequences and z for corresponding labels. Both are compared with the edit distance function $ED(\cdot, \cdot)$ which is the number of characters that must be changed to make the input sequence equal to the corresponding label. The Sequence Error Rate (SER) is the percentage of predictions in the test set that are **not completely correct**.

You can see example images of a random character sequences after the different preprocessing steps in Table 3. For every blur radius, a training set is created based on the different

Radius	LER			SER		
	FPP	NAB	NB	NAB	FPP	NB
2	17.73 %	29.99 %	34.20 %	64.06 %	53.91 %	84.38 %
3	19.98 %	88.66 %	25.90 %	100.00 %	62.11 %	71.48 %
4	12.22 %	55.52 %	28.54 %	94.92 %	46.09 %	64.84 %
5	18.24 %	20.04 %	36.36 %	63.28 %	53.13 %	72.27 %
6	33.60 %	49.20 %	32.54 %	86.72 %	73.44 %	75.39 %
7	51.95 %	34.27 %	28.67 %	76.17 %	92.58 %	66.02 %
8	28.97 %	79.24 %	61.22 %	100.00 %	68.75 %	95.70 %
9	60.50 %	31.35 %	46.92 %	76.17 %	96.48 %	90.23 %
10	50.97 %	74.88 %	37.17 %	100.00 %	94.53 %	83.98 %
11	47.88 %	37.76 %	43.09 %	82.42 %	94.14 %	89.06 %
12	56.12 %	52.06 %	61.29 %	96.88 %	97.66 %	97.66 %
13	55.20 %	61.38 %	53.34 %	100.00 %	94.92 %	92.97 %
14	54.27 %	42.06 %	59.01 %	87.89 %	95.70 %	98.05 %
15	58.33 %	34.63 %	53.93 %	81.64 %	98.05 %	96.48 %
16	40.41 %	56.12 %	47.62 %	95.31 %	83.98 %	90.23 %
20	46.74 %	55.65 %	54.21 %	96.48 %	88.28 %	97.66 %
24	55.72 %	57.69 %	50.67 %	98.44 %	98.83 %	95.70 %
32	57.12 %	62.16 %	63.04 %	98.44 %	98.83 %	100.00 %

Tab. 1: Error rate of the proposed depixelation system, see Fig. 2 and Fig. 3 for chart illustration and Tab. 3 for training set examples.

JPEG Quality Level	**LER**	**SER**
100	32.84%	75.00%
50	29.73%	69.14%
10	38.38%	81.64%

Tab. 2: JPEG Quality Loss Experiment with a blurring radius of 10×10 pixels

configurations described in subsection 4.2. Note that the scale of the input image does not change. This becomes clear in the NAB column where no alignment is performed. When increasing the blur radius, the blurred area takes more space on the canvas, especially towards the borders. The visualization component (preview) of our preprocessors with activated alignment scales the images up and shows a centered view so that it looks like the images would have different lengths which is not the fact.

6 Discussion

In our experiment, we analyze whether a Machine Learning model and corresponding pipeline for retrieving data from redacted images can be used for blurred images as well. The paper from Garske and Noack [GN22] demonstrates this on pixelized images only.

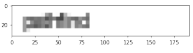
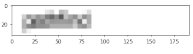
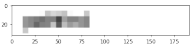
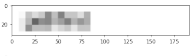
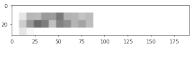
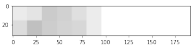
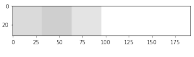
Radius	FPP	NAB	NB	Best LER
2				17.73 % (FPP)
3				19.98 % (FPP)
4				12.22 % (FPP)
5				18.24 % (FPP)
6				32.54 % (NB)
7				28.67 % (NB)
8				28.97 % (FPP)
9				31.35 % (NAB)
10				37.17 % (NB)
11				37.76 % (NAB)
12				52.06 % (NAB)
13				53.34 % (NB)
14				42.06 % (NAB)
15				34.63 % (NAB)
16				40.41 % (FPP)
20				46.74 % (FPP)
24				50.67 % (NB)
32				57.12 % (FPP)

Tab. 3: Examples for each blur radius in the different test passes

Our test results indicate that it is actually possible to train such a network on blurred images. Naturally, when the blur radius rises, the LER and SER will rise, too. However, the network can still recover information even if a human has trouble to decipher the image. We assume a blur radius of 10 pixels as the limit up to where a human can still recognize all characters without problems.

When considering pixelation and blurring, the question arises how they relate to each other. Tab. 4 provides a visual aid for comparing some block sizes of pixelized texts and blur radii of blurred texts. It becomes clear that users need a higher blur radius in order to achieve a comparable visual redaction level as with pixelation.

When comparing the LERs for both anti-redaction techniques (e.g. 32x32 px), we can see that our initial working hypothesis is confirmed. Retrieving data from a blurred text works at least as accurate as for pixelized text. That means blurring has in the best case the same security as pixelation, rather a little bit less. Note that [GN22] use a slightly different dataset than it is used in this experiment: for performance reasons, only 40,000 instead of 100,000 unique passwords are used as a wordlist for the training. Additionally, we anticipate that each model is trained for those parameters (font size, etc.) that are also used for the test images.

Block Size	Example Image	LER	Blur Radius	Example Image	LER
2×2		2.99 %	2×2		17.73 %
3×3		16.00 %	5×5		18.24 %
4×4		20.47 %	7×7		28.67 %
5×5		36.17 %	9×9		31.35 %
6×6		28.92 %	10×10		37.17 %
7×7		41.17 %	12×12		52.06 %
8×8		49.93 %	14×14		42.06 %
16×16		61.09 %	20×20		46.74 %
32×32		69.25 %	32×32		57.12 %

Tab. 4: Visual aid for the comparison of pixelized and blurred images. Block sizes and blur radii are optically fitted. The examples for pixelized images are taken from [GN22].

7 Conclusion and Future Work

In this paper, we analyze the privacy of blurring whereby our research especially focuses on the obfuscation of credentials like passwords. We show that a neural networks can not only be used for recovering information from pixelized images but also from blurred images.

We point out that higher blur radii are required to enable comparable results regarding pixelation. If investigators are able to guess parameters like font family or font size, they can run a dictionary-like attack to create a model that can reconstruct redacted parts in e.g. publicly available images. Besides, it turned out that our pipeline and model are able to handle JPEG quality loss with only slightly increased error rates.

There is still room for improvement: the current architecture and preprocessor can be further examined to improve the LER and SER metrics and thus the results. In addition to it, the influence of other parameters like different font families can be investigated. Moreover, aspects of transfer learning remain open: the model can be extended in a way to transfer the knowledge from specific parameters to applications with more generic parameters. By doing so, one model could support different font sizes or even a variety of redaction techniques without the need for retraining every time.

References

[CFV11] Cavedon, L.; Foschini, L.; Vigna, G.: Getting the face behind the squares: Reconstructing pixelized video streams. In: 5th USENIX Workshop on Offensive Technologies (WOOT 11). 2011.

[GN22] Garske, V.; Noack, A.: Recovering information from pixelized credentials. In: *Lecture Notes in Informatics, Proceedings*, Sicherheit 2022, Apr. 8, 2022. Karlsruhe, pp. 129–141, Apr. 2022, ISBN: 978-3-88579-717-3.

[Gr06] Graves, A.; Fernández, S.; Gomez, F.; Schmidhuber, J.: Connectionist Temporal Classification: Labelling Unsegmented Sequence Data with Recurrent Neural Networks. In: Proceedings of the 23rd International Conference on Machine Learning. ICML '06, Association for Computing Machinery, Pittsburgh, Pennsylvania, USA, pp. 369–376, 2006, ISBN: 1595933832, URL: https://doi.org/10.1145/1143844.1143891.

[Hi16] Hill, S.; Zhou, Z.; Saul, L.; Shacham, H.: On the (in) effectiveness of mosaicing and blurring as tools for document redaction. Proceedings on Privacy Enhancing Technologies 2016/4, pp. 403–417, 2016.

[Ka21] Kaiser, P.; Schirrmacher, F.; Lorch, B.; Rieß, C.: Learning to Decipher License Plates in Severely Degraded Images. In: MultiMedia FORensics in the WILD - accepted for publication. Jan. 11–11, 2021, URL: https://faui1-files.cs.fau.de/public/publications/mmsec/2021-Schirrmacher-LTD.pdf.

[LCC21] Lundh, F.; Clark, A.; Contributors: Pillow: The friendly PIL fork (Python Imaging Library) repository, Version 8.3.1, Apr. 1, 2021, URL: https://github.com/python-pillow/Pillow/tree/8.3.1.

[Me20] Menon, S.; Damian, A.; Hu, S.; Ravi, N.; Rudin, C.: Pulse: Self-supervised photo upsampling via latent space exploration of generative models. In: Proceedings of the ieee/cvf conference on computer vision and pattern recognition. Pp. 2437–2445, 2020.

[MHg21] Miessler, D.; Haddix, J.; g0tmi1k: SecLists, Mar. 1, 2021, URL: https://github.com/danielmiessler/SecLists.

[MSS16] McPherson, R.; Shokri, R.; Shmatikov, V.: Defeating Image Obfuscation with Deep Learning, 2016, arXiv: 1609.00408 [cs.CR].

[Ni19] Nischwitz, A.; Fischer, M.; Haberäcker, P.; Socher, G.: Bildverarbeitung: Band II des Standardwerks Computergrafik und Bildverarbeitung. Springer-Verlag, 2019, ISBN: 978-3-658-28704-7.

[SRP19] Soullard, Y.; Ruffino, C.; Paquet, T.: CTCModel: a Keras Model for Connectionist Temporal Classification, 2019, arXiv: 1901.07957 [cs.LG].

[Xu17] Xu, X.; Sun, D.; Pan, J.; Zhang, Y.; Pfister, H.; Yang, M.-H.: Learning to super-resolve blurry face and text images. In: Proceedings of the IEEE international conference on computer vision. Pp. 251–260, 2017.

A Additional Information

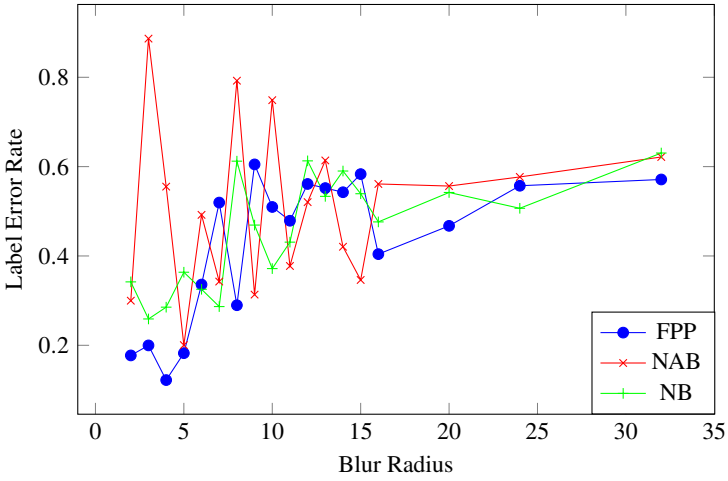


Fig. 2: Label error rates (LER) in relation to the blurring radius

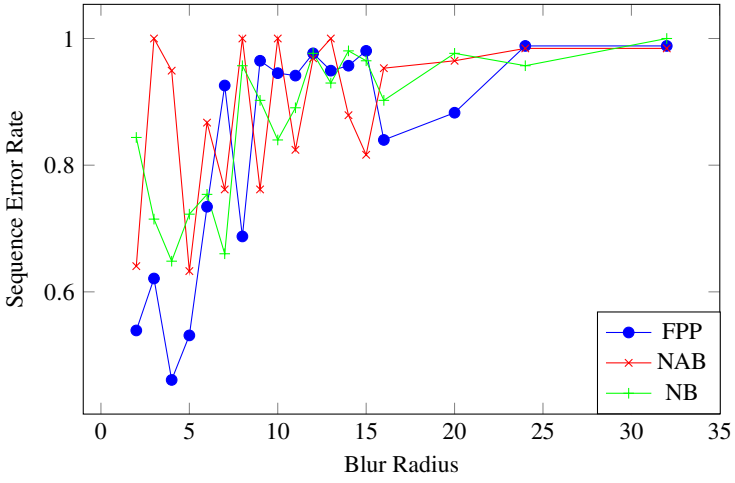


Fig. 3: Sequence error rates (SER) in relation to the blurring radius

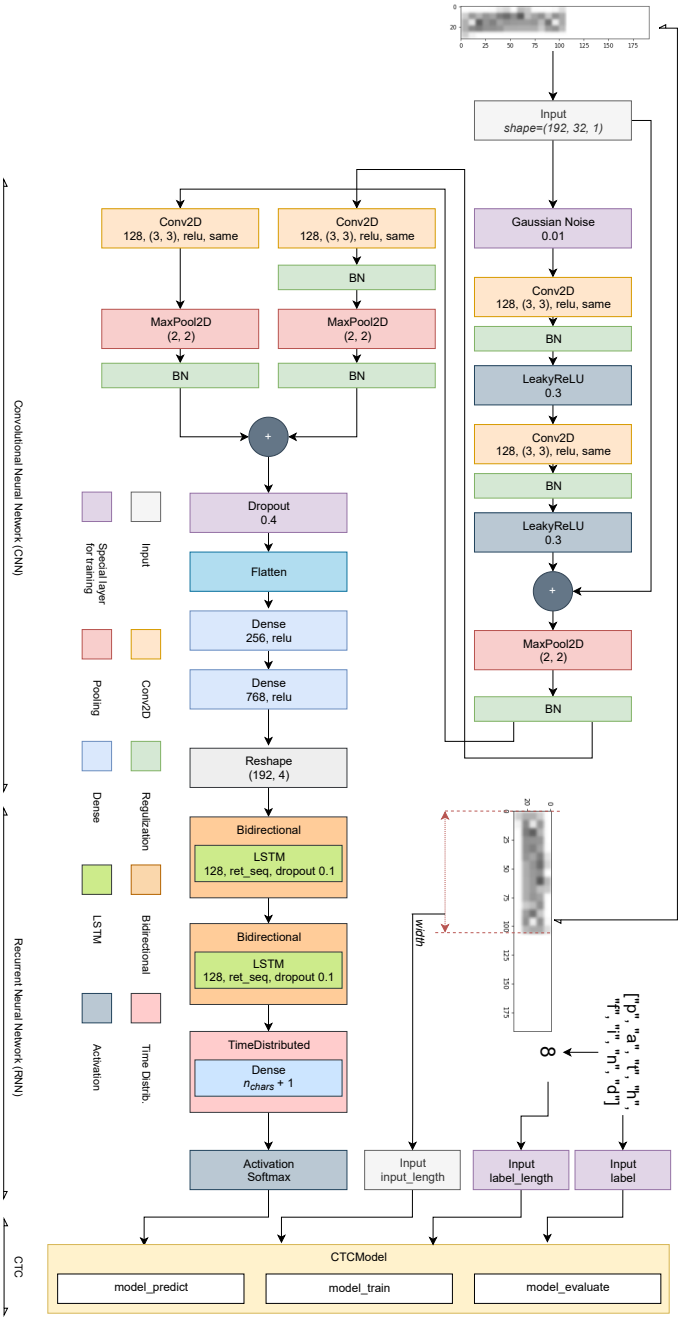


Fig. 4: Architecture of the neural network [GN22]

Ontology in the Digital Forensics Domain: A Scoping Review

Martin Morgenstern[1], Johannes Fähndrich[2] and Wilfried Honekamp[3]

Abstract: A need for the use of automation in digital forensics is imminent. With the overwhelming workload for analysing collected digital evidence, law enforcement agencies are no longer meeting the quality investigations we expect. For automation to work (e. g., integrating heterogeneous data sources, structuring unstructured data, or drawing conclusions from structured data), formalisation must be specified. Specification of formalisation includes as a first step to describe concepts in the domain of digital forensics. In this contribution, we analyse the state of the art of ontological formalisation in the domain of digital forensics via a scoping review. There are some attempts to formalise the technical domain of digital forensics in ontologies, but they do not cover essential context, like technical annotations or acquisition modelling. Future work will be to use the existing foundations and extend them with facts and rules to enable reasoning.

Keywords: Formalisation, Digital Forensics, Ontology.

1 Introduction and Background

When collecting evidence in criminal cases, digital traces must be identified, recorded and evaluated. Due to the quantity and heterogeneity (see Figure 1) of forensically relevant digital data, it is no longer feasible for investigating authorities to perform a high-quality manual analysis of the same. In order to automate the analysis of extensive and heterogeneous data in a meaningful way, for example by using artificial intelligence, a generally applicable ontology is needed for the field of digital forensics. This paper aims to highlight the current state of development on the topic of ontologies for general IT forensics, as well as the resulting research gaps to provide a basis for the development of forensic AI systems. The overarching goal is a proposal for a uniform use of technical language. Therefore, it will be investigated which ontological prerequisites have to be created in IT forensics to enable a meaningful use of artificial intelligence for IT forensic analysis, and it will be reviewed, which of these prerequisites are already fulfilled.
In order to solve crimes, facts have to be found out and documented in a way that can be used in court. Very often, this requires the use of forensic experts. Forensics is the science that deals with the identification, analysis and reconstruction of criminal acts.

[1] Hochschule Stralsund, IACS, Zur Schwedenschanze 15, Stralsund, 18435, martin.morgenstern@hochschule-stralsund.de
[2] Hochschule für Polizei Baden-Württemberg, Johannesfaehndrich@hfpol-bw.de
[3] Hochschule Stralsund, IACS, Zur Schwedenschanze 15, Stralsund, 18435, wilfried.honekamp@hochschule-stralsund.de, https://orcid.org/0000-0003-2931-7047

doi:10.18420/inf2022_05

Forensics is divided into several subfields. One subfield is computer forensics, which is a synonym for digital forensics [Ch09]. [Si18]

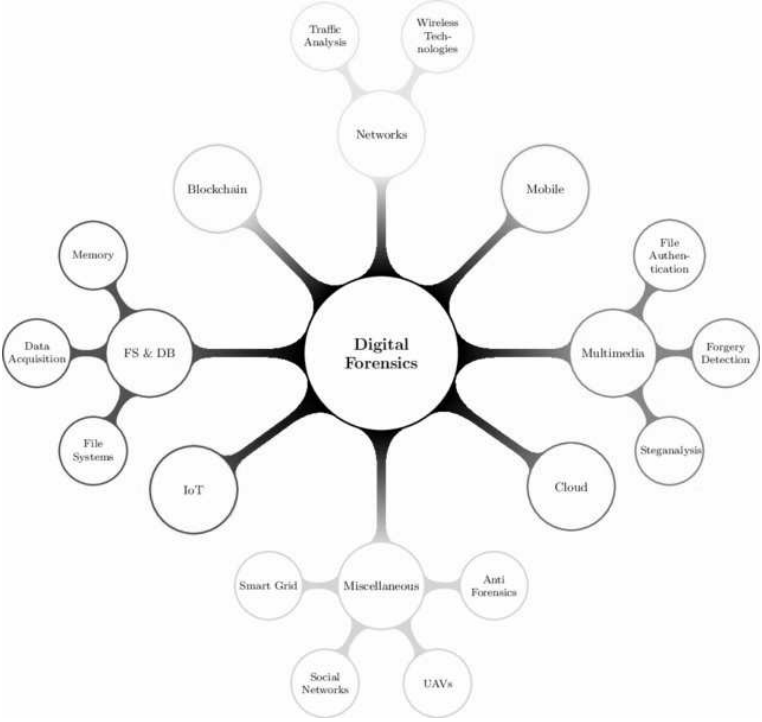


Figure 1: Sub-areas of digital forensics [Ca21]

Figure 1 shows an overview of identified sub-areas of digital forensics. However, the sub-areas can only ever be a snapshot, as information technology is constantly changing and evolving. The authors consider it problematic that they are classified according to data origin. Povalej et al. already drew attention to the problem in 2021 and asked whether such a division is still up-to-date or whether the classification should be done according to the type of digital traces. [Po21].

Another term for digital forensics is IT forensics. In this paper, the definition of digital forensics used is that of the German Federal Office for Information Security (BSI). According to this definition, digital forensics is "the strictly methodical analysis of data on data carriers and in computer networks to clarify incidents, including the possibilities of strategic preparation, especially from the point of view of the system operator of an IT system." [BS11]

The complexity and quantity of digital forensics is increasing every year. On the one hand, the amount of data is permanently increasing, and on the other hand, so is the number of possible data sources. One reason for this are rapidly falling prices for large data media. [Al13] In traditional IT forensics, data carriers are removed, copied for forensic purposes and analysed. Supporting tools are used for the evaluation. The use of

artificial intelligence could significantly facilitate IT forensic work. In addition to legal hurdles, the lack of ontological integration of heterogeneous data is also an obstacle to the use of AI in digital investigations [Po21], [Ho22]. Errors in the evaluation of digital traces can have fatal, e. g. financial or social, consequences. One cause of evaluation errors can be ignorance of exactly how forensic evaluation software works or the assumption that the software always works without errors. In 2016, for example, a suspect was nearly acquitted because the forensic evaluation software failed to process a memory area on the smartphone being evaluated. In individual areas of computer forensics, e. g. mobile forensics, cooperation with external service providers is necessary. For the cooperation with external service providers, the standardisation of forensic procedures is useful. [He21]

For successful cooperation between different organisations, in addition to the standardisation of processes, a common description of the same is also necessary. However, in the field of digital forensics, there is a lack of a generally accepted general ontology [Gr17]. Ontologies are used to represent and provide a basis for logical reasoning in a domain, or for objectifying knowledge (for example, in a knowledge graph [So99]). There are several definitions for the term ontology. [Ka14], [He05].

According to Gruber's definition, one of the most well-known definitions of ontology, an ontology is a common specification to share concepts. Through an ontology a knowledge domain, a domain, is described. Another definition of ontology that may be helpful here is "An ontology is a formal, explicit specification of a shared conceptualisation." [St98]. In this description, a standardised terminology is used, and relationships and derived rules are defined. In many knowledge domains there are multiple or competing ontologies. An ontology gains value the more it is recognised in the professional community. [He05]

For example, a technical application of ontologies is illustrated by Linked Open Data [Mc20]. Currently, while there are many digital forensics ontologies, none are widely used. In the past, there have been several approaches for the introduction of a general ontology in digital forensics, but these do not yet meet the current requirements. The requirements for an ontology for digital forensics differ depending on the literature. From the literature research some requirements emerge, which can be found again and again. In particular, ontologies should be able to describe data from heterogeneous sources. In a modern ontology for digital forensics, cross-agency and cross-national work should also be possible. The authors assume that a common and standardised ontology will improve communication between investigative and law enforcement agencies and, if necessary, external forensics experts, thus contributing to an increase in efficiency in law enforcement. Ontologies are used to support the investigation of data from different types of sources in various application areas of digital forensics [Av10]. One of the main challenges is to find a way to automatically create structured data from different data sources [Si21]. The need for formalisation in digital forensics was noted as early as 2004 by Bruschi, Monga, and Marignoni. At the digital forensic Research Conference, a hypothesis-based scheme for formalising contexts in IT forensic analysis was presented. However, no ontology was developed in this work. [Br04]

Model driven engineering (MDE) abstracts models domains and implementations for formalisation and normalisation [Fa04]. With MDE one idea is to domain-specific

modelling language to aid “formalisation the application structure, behaviour and requirements” [Sc06]. MDE enables the description of self-adapting systems [Tr18]. For self-adaptation to be useful, the semantics of the models used can support the adaptation process. Based on this idea of MDE, semantic service description can be seen as a machine-readable model including meta-models, transformations and semantics. An ontology formalizes the concepts used in a domain to describe the parameters and results of services. „High quality domain ontologies are essential for successful employment of semantic Web services“ [Sa05] (p. 1).

The idea of this formalisation is to formally extract shared knowledge in order to facilitate a common explicit understanding of the field of digital forensics. The idea is to create a common language, structured documentation and tools, which stand up in curt to structure forensic work. Much of the knowledge about digital forensics is implicit in experts, and is built upon experience. The aim is to share this knowledge. First formation of a vocabulary is part of the formalisation with an ontology. We see this as an extension of the Linked Open Data, where knowledge of different domains has been collected and formalised for others to use.

Similar to Euzenat and Shavaiko [Eu07] we define an Ontology the following way:

$$\mathbf{o} = \{C, I, R, \leq, \perp, \in, =\}$$

With C is the set of classes, I is the set of Individuals, R is the set of relationships, $\leq$ is a relation on $(C \times C) \cup (R \times R) \cup (T \times T)$ called specialization, $\perp$ is a relation on $(C \times C) \cup (R \times R) \cup (T \times T)$ called exclusion, $\in$ is a relation over $((I \times C) \cup (V \times T)$ called instantiation, = is a relation over $I \times (I \cup V)$ called assignment.

Euzenat and Shavaiko [Eu07] (p. 39), define that expression can be done and are satisfied by an interpretation if it is coherent. We then are able to create formulars over this ontology with e.g., an implementation like Web Ontology Language (OWL): This gives us basic entities, which can have relation with other entities or themselves. We will use this formal definition to ground further work in OWL, regarding Resource Description Framework (RDF) triplets and the formulation of entities and relations as well as the instantiation by e.g., datatypes and Individuals. Thus, we will use this kind of ontology to state state facts (Individuals and their relationships) about entities (TBox) and facts about the abstract classes (ABox) (concepts and their relationships). This can be seen as a formal grounding of Cyber-investigation Analysis Standard Expression (CASE) or Unified Cyber Ontology (UCO) and the reuse of this first formalization. The benefits of formalising digital forensics were presented in detail in Dewald's dissertation. In this, digital traces were formalised based on an abstract model. Through the formalisation it was possible to show if and which relationship exists between different traces. [De12]

2 Method

To capture the current state of development of ontologies in digital forensics, a scoping review was conducted. A scoping review is a suitable method to obtain an overview of

the evidence of a research topic. Scoping review can identify research gaps in the area under investigation. [Vo19]
Alternativ zum Scoping Review gibt es weitere wissenschaftliche Methoden um einen Überblick über die relevante Literatur, sowie dem Stand der Forschung zu erhalten. Alternativ hätte z. B. eine strukturierte Literaturrecherche gewählt werden können. Der Vorteil des Scoping Reviews ist die hohe Effienz, da mit wenig Zeitaufwand aktuelle wissenschaftliche Fragestellungen identifiziert werden können. [Vo19]
Only German and English language sources were reviewed for the scoping review. Google and Google Scholar were used to search for relevant literature. Search terms were 'ontology digital forensics' as well as synonyms in German and English. For both Google and Google Scholar, only sources on the first five pages of search results were used. Furthermore, the Google Scholar function "cited by" was used to find further relevant documents. All searches took place in May 2022.

3 Results

In the following, the results of the literature review from 2009 until today are presented chronologically. The need for ontology for digital evidence analysis was identified at least in 2009, when Kahvedžić and Kechadi described the DIALOG framework for modelling knowledge in the field of digital forensics (see Figure 2). For a detailed description, the sub-ontologies criminal case, information, information location, and forensic resource were used. The example of the Windows Registry was used to illustrate the use of DIALOG. digital forensics is the central element of the ontology. The first level of sub ontologies are criminal case, information, location information and forensic recovery. The authors of DIALOG have indicated in their work that further work will be needed to consider additional use cases, such as file system analysis or mobile phone forensics. [Ka09]

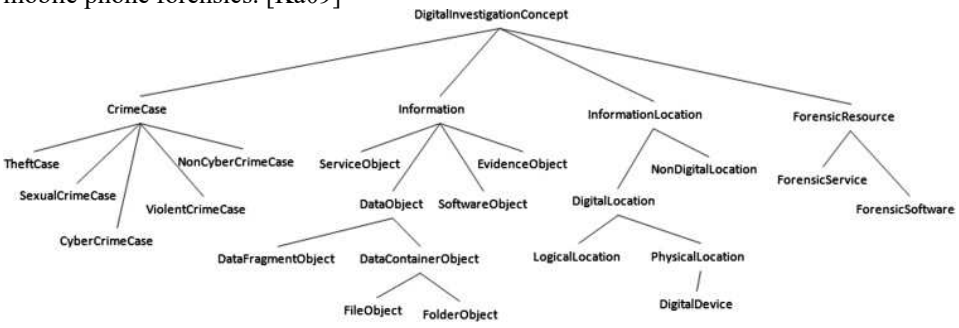


Figure 2: Hierarchy of the DIALOG framework [Ka09]

Back in 2011, Poisel and Tjoa noted that there has been progress in the area of IT forensics ontology, but its unification of standardisation is the biggest challenge. [Po11] Especially the comparability of created reports was seen by Bariki, Hashmi, and Baggili as a problem to be solved. Therefore, an XML-based proposed solution was published.

The development or extension of a standard that also includes other areas such as memory or network forensics was seen as a future challenge. [Ba11] Similarly, in 2012, a conference paper from n Ćosić and Ćosić detailed the need for a unified ontology for digital forensics [Co12].

In 2013, a framework for IT forensic analysis based on semantic web technologies was presented that was based on an ontology. Although semantic ontologies for the digital forensics domain already existed at that time, according to the authors, they focused only on evidence management and forensic reporting. However, the focus of the new framework was on digital evidence analysis. The goal was to develop a framework that would make it possible to link information from different sources in a way that would result in new information. [Al13]

In 2015, Casey et al found that there was no widely accepted standard representation in the digital forensics domain. The existing Structed Thread Information eXpression (Stix) and other models for exchanging legally relevant information could not specifically address the specifics of the digital forensics domain. As a result, it would not be possible to compare investigative results across tools and agencies. Various standards, such as digital forensic XML, have been developed to enable the exchange of IT forensic information. However, none of these standards has gained widespread acceptance. The development of the digital forensic Analysis eXpression (DFAX) ontology was intended to solve the problems described above. Included in DFAX were many improvements over previous standards, such as including chain of custody evidence and victim and investigator actions [Ca15].

Ćosić and Baca developed the Digital Evidence Framework in 2015. The Developed Ontology was intended to solve recognised problems of previous ontologies. One of the innovations was that the integrity of digital evidence should also be considered in the ontology. Lack of standards and processes for sharing information between different agencies and states was identified as a problem to be solved [Co15].

As of October 2016, DFAX is no longer under development. Since that time, the project page states that information about a new project under development will be linked there. [Ba16] Even for basic terms like artifact, there were different definitions. This problem also highlighted the need for a universally accepted and recognised ontology for digital forensics. Harichandran et al. sought a formal definition for digital forensic artifacts [Ha16].

There was still a need for a standard to represent digital evidence and its interrelationships. However, DFAX did not meet all the requirements. For example, it was not possible to represent deleted data in this schema. A new standard was developed in the form of the Cyber-investigation Analysis Standard (CASE). The development of the new standard incorporated experience from many previous standards, including DFXML and DFAX. Improvements were made for national and international exchange of information [Ca17].

Semantic and ontological linking of forensically relevant information has received increasing attention from scholars in recent years [Bh20]. For example, in 2019, Amato et al. presented a semantic methodology based on Natural Language Processing (NLP) that used log data as an example to represent unstructured data in a standardised format. [Am20] OliveiraJr et al. published a paper on experimental digital forensics. The goal

was to find a method to describe experiments in the domain of digital forensics in a repeatable way. The development of an ontology for experimental digital forensics was identified as a future task. [Ol20]
In digital forensics today, a variety of different data sources and types are evaluated. In addition to traditional data carriers, online profiles, cloud data, IoT devices with different operating systems and file systems, smartphones, and other devices often need to be evaluated today. In addition to the quantity of data and devices to be evaluated, the complexity is also increasing, for example due to encryption. For many areas of digital forensics, there were 2020 ontologies, e.g., for social networks or web services forensics. While technical capabilities for analysing data continue to evolve, there is little research in digital forensics ontology. There is a need for standardisation of digital forensics ontologies, as existing ontologies are not widely used or belong to subset of digital forensics, e.g., cloud forensics ontologies [Ke20]. The ontology to be developed should be developed globally and have wide user acceptance. Standardisation of data is necessary to process it in an automated way, such as rule-based or machine learning [Si21].
Recently, Solanke and Biasiotti published a paper on Formalisation of AI in digital forensics. In it, they presented techniques for analysing the effectiveness of classification and regression in digital forensics. Further development of methods for evaluating AI techniques in digital forensics is seen as a promising development [So22].

4 Discussion and Conclusions

There are first attempts to formalise the technical domain of digital forensics in an ontology. However, only CASE as an extension of UCO has a description logic class of ALUO(D). Here Object-Properties, data properties and individual types are defined. As a formalisation, this means that a solid foundation has been established. However, in order to make such an ontology usable for automation, for example, technical annotations or acquisition modelling is still needed. For example, an IP address is modelled, but not yet what type of IP exists on a system (e.g., public or private) and where that IP appears in an investigation, for example, the IP of the victim or the perpetrator. Another annotation missing is the tools of the operating systems by means of which an IP can be collected, documented, and used in an investigation. So, for the example IP, ipconfig or ifconfig should be described for the local view, nslookup for the resolution, ping or tracert/traceroute for the classification in a network and so on. Criticism of the previous approaches is that mostly only the classification (like the introduction of IPv4 and IPv6 as a subclass of the IP address) has been modelled here, but hardly any further forensic knowledge has been formalised. The technical knowledge that an IPv4 is also an IP address is basic but not yet forensic. What is important for an IP address in an investigation (deletion periods, resolution at ISP, individualisation through assignment to MAC address, etc.) is missing here. We implemented an example of the CASE and UCO ontology to show how the implementation of OWL Object-Properties and Individuals can be used to describe knowledge of the utilisation of tools, their parameters, and the

interrelation of the described entities in the ontology. This has the goal to automate the appropriate tasks to make the work in digital forensics manageable in the future. We used an example of an IP address, a MAC address, its assignment to the network interface, and the appropriate tools.

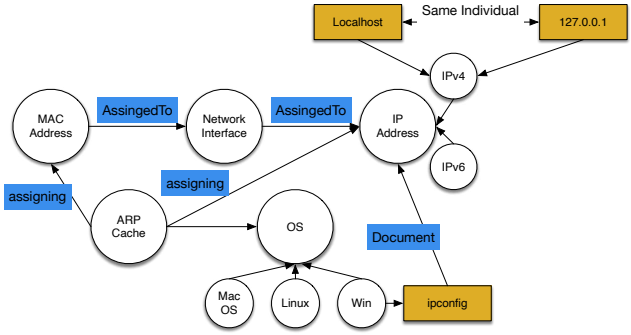


Figure 3: Extension of the CASE ontology by example Individuals and Object-Properties.

Figure 3 shows how the relationship to tools like ipconfig and the use of Object-Properties to externalise the knowledge that the ARP cache holds information about which IP address has been assigned to which MAC address. The aim is to use reasoners and semantic service descriptions to automate some of the information extraction processes in digital forensics cases.

So future work will be to use the existing basis of UCO and CASE and extend it with rules of the Semantic Web Rule Language, and for example Linear Time Logic, to be able to formalise facts like "IP addresses are collected by the ISP until the deletion period of data retention is reached and therefore the connection owner must be documented with a request for information until then". Reasoning could then take place on this, which, for example, sends requests independently or prioritises for which trace situation which necessary next steps must be initiated.

Bibliography

[AL13] Alzaabi, M., Jones, A., Martin, T.A.: An Ontology-Based Forensic Analysis Tool. ADFSL Conference on Digital Forensics, Security and Law, 2013.

[Am20] Amato, F., Castiglione, A., Cozzolino, G., Narducci, F.: A semantic-based methodology for digital forensics analysis. J. Parallel Distrib. Comput. 138, 172–177, 2020.

[Av10] Avni, O., Knierim, T.: Carving und semantische Analyse in der digitalen Forensik. Frauenhofer IGD-A8 Sicherheitstechnologie. 2010.

[Ba16] Back, G.: dfax (DEPRECATED). DFAX. 2016. github.com/DFAX/dfax (accessed 5.23.22).

[Ba11] Bariki, H., Hashmi, M., Baggili, I.: Defining a Standard for Reporting Digital Evidence Items in Computer Forensic Tools. In: Baggili, I. (eds) Digital Forensics and Cyber Crime. ICDF2C 2010. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 53. Springer, Berlin, Heidelberg. 2011.

[Bh20] Bhandari, S., Jusas, V.: An Ontology Based on the Timeline of Log2timeline and Psort Using Abstraction Approach in Digital Forensics. Symmetry 12, 642. 2020.

[Br04] Bruschi, D., Mongra, M., Martignoni, L.: How to Reuse Knowledge about Forensic Investigations, The Digital Forensic Research Conference. Baltimore. 2004.

[BS11] BSI, 2011. Leitfaden IT-Forensik. Bundesamt für Sicherheit in der Informationstechnik. Bonn. 2011

[Ca15] Casey, E., Back, G., Barnum, S.: Leveraging CybOXTM to standardize representation and exchange of digital forensic information. Digit. Investig. 12, pp. 102-110. 2015.

[Ca17] Casey, E., Biasiotti, M.A., Turchi, F.: Using Standardization and Ontology to Enhance Data Protection and Intelligent Analysis of Electronic Evidence. 2017.

[Ca21] Casino, F., Dasaklis, T., Spathoulas, G., Anagnostopoulos, M., Ghosal, A., Borocz, I., Solanas, A., Conti, M., Patsakis, C.: A cross-domain qualitative meta-analysis of digital forensics: Research trends, challenges, and emerging topics., 2021.

[Ch09] Chu, H.-C., Deng, D.-J., Chao, H.-C.: An Ontology-driven Model for Digital Forensics Investigations of Computer Incidents under the Ubiquitous Computing Environments. Wirel. Pers. Commun. 56, pp. 5-19. 2009.

[Co15] Ćosić, J., Baca, M.: Leveraging DEMF to Ensure and Represent 5ws&1h in Digital Forensic Domain 13, 5. 2015.

[De12] Dewald, A.: Formalisierung digitaler Spuren und ihre Einbettung in die Digitale Forensik. Universität Erlangen-Nürnberg. 2012.

[Eu07] Euzenat, J., & Shvaiko, P. (2007). Ontology matching (Vol. 18). Heidelberg: springer.

[Fa04] Favre, J. M. (2004, October). Towards a basic theory to model model driven engineering. In 3rd workshop in software model engineering, wisme (pp. 262-271).

[Gr17] Grigaliunas, S., Toldinas, J., Venckauskas, A.: An Ontology-Based Transformation Model for the Digital Forensics Domain. Elektron. Ir Elektrotechnika 23, pp. 78-82. 2017

[Ha16] Harichandran, V.S., Walnycky, D., Baggili, I., Breitinger, F.: CuFA: A more formal definition for digital forensic artifacts. Digit. Investig. 18, p. 125-137. 2016.

[He21] Heller, P.: Probleme bei digitalen Ermittlungen - Wenn Forensik-Software zu Fehlurteilen führt. Deutschlandfunk. 2021. URL www.deutschlandfunk.de/probleme-bei-digitalen-ermittlungen-wenn-forensik-software-100.html (accessed 5.27.22).

[He05] Hesse, W.: Ontologie(n). 2005. URL gi.de/informatiklexikon/ontologien (accessed 5.24.22).

[Ho22] Honekamp, W., Povalej, R., Rittelmeier, H., Fähndrich, J., Berner, S., Labudde, D.: Technologiegetriebene Polizeiausbildung im Umgang mit Digitalen Spuren, in: Rüdiger, T.-G., Bayerl, P. S. (Eds.) Handbuch Cyberkriminologie. Springer. 2022.

[Ka09] Kahvedžić, D., Kechadi, T.: DIALOG: A framework for modeling, analysis and reuse of digital forensic knowledge. Digit. Investig. 6, pp. 23-33. 2009.

[Ka14] Karie, N.M., Venter, H.S.: Toward a General Ontology for Digital Forensic Disciplines. J. Forensic Sci. 59, pp. 1231-1241. 2014.

[Ke20] Kebande, V.R., Karie, N.M., Ikuesan, R.A., Venter, H.S.: Ontology-driven perspective of CFRaaS. WIREs Forensic Sci. 2, e1372. 2020.

[Mc20] McCrae, J. P.: The Linked Open Data Cloud. 2020. URL lod-cloud.net (accessed 5.27.22).

[Ol20] OliveiraJr, E., Zorzo, A.F., Neu, C.V.: Towards a conceptual model for promoting digital forensics experiments. Forensic Sci. Int. Digit. Investig. 35, 301014. 2020

[Po11] Poisel, R., Tjoa, S.: Forensische Untersuchung multimedialer Daten, in: 9. Sicherheitskonferenz Krems. Donau-Universität Krems, 2011.

[Po21] Povalej, R., Rittelmeier, H., Fähndrich, J., Berner, S., Honekamp, W., Labudde, D.: Die Enkel von Locard: Analyse digitaler Spuren in der forensischen Informatik. Inform. Spektrum 44, pp. 355–363. 2021

[Sa05] Sabou, M., Wroe, C., Goble, C., & Stuckenschmidt, H. (2005). Learning domain ontologies for semantic web service descriptions. Journal of Web Semantics, 3(4), 340-365.

[Sc06] Schmidt, D. C. (2006). Guest editor's introduction: Model-driven engineering. Computer, 39(02), 25-31.

[Si21] Sikos, L.F.: AI in digital forensics: Ontology engineering for cybercrime investigations. WIREs Forensic Sci. 3. 2021

[Si18] Siller, H.: Definition: Forensik. 2018. URL wirtschaftslexikon. gabler.de/definition/forensik-53390/version-276483 (accessed 5.27.22).

[So22] Solanke, A.A., Biasiotti, M.A.: Digital Forensics AI: Evaluating, Standardizing and Optimizing Digital Evidence Mining Techniques. KI - Künstl. Intell. 2022.

[So99] Sowa, J.F.: Knowledge representation: logical, philosophical, and computational foundations. Brooks/Cole, Pacific Grove. 1999.

[St98] Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: Principles and methods. Data Knowl. Eng. 25, pp. 161–197. 1998.

[Tr18] Trollmann, F., Fähndrich, J., & Albayrak, S. (2018, May). Hybrid adaptation policies: towards a framework for classification and modelling of different combinations of adaptation policies. In Proceedings of the 13th International Conference on Software Engineering for Adaptive and Self-Managing Systems (pp. 76-86).

[Vo19] von Elm, E., Schreiber, G., Haupt, C.C.: Methodische Anleitung für Scoping Reviews (JBI-Methodologie). Z. Evid. Fortbild. Qual. Gesundh. wesen 143. 2019.

SCALA-Speech: An Interactive System for Finding and Analyzing Speech Content in Audio Data

Alessia Cornaggia-Urrigshardt[1] Nikita Jarocky[1] Frank Kurth[1] Sebastian Urrigshardt[1] Kevin Wilkinghoff[1]

Abstract: Audio data does not contain as much static information as images and texts and thus analyses inherently require more time. Although in monitoring applications it is likely that large quantities of the captured audio files do not contain meaningful information, without prior knowledge investigators need to listen to all audio files in full length. In this work, a system for automatically finding and analyzing speech content in audio data is presented. The system provides different speech processing algorithms as well as a graphical interface (SCALA) for assisting investigators in audio analysis. The system consists of four components: speech detection, language recognition, speaker diarization/recognition and keyword spotting. SCALA-Speech structures audio data by recognizing speech regions, used languages and speaker changes, thus enabling investigators to listen to audio data more efficiently. Furthermore, specific speakers and keywords can be annotated and searched for. Usage of SCALA-Speech is demonstrated on audio tracks of videos linked in Twitter posts related to an exemplary topic.

Keywords: Audio Monitoring; Speech Detection; Language Recognition; Speaker Diarization; Keyword Spotting; Deep Learning

1 Introduction

The general goal of processing digital data in law enforcement is to gather information about potential crimes or suspects and collect potential evidence. However, analyzing large amounts of digital data requires substantial amounts of manpower and time although most data files or large parts of them do not contain any interesting information. Analyzing audio data such as speech is even more costly because it does not contain static information and thus listening to it requires time proportional to its length. Additionally, there are several difficulties specific to speech data: First, multiple languages can be encountered in which a particular investigator may lack proficiency. Furthermore, it is difficult for a human to remember and correctly recognize a larger set of unfamiliar speakers. Hence, an interactive system for assisting investigators in analyzing audio data is vital.
There are several use cases for processing audio data in law enforcement. One prominent example is wire-tapping telephone lines [Di21] and more generally fighting organized

[1] Fraunhofer Institute for Communication, Information Processing and Ergonomics FKIE, Fraunhoferstraße 20, 53343 Wachtberg, Germany, alessia.cornaggia-urrigshardt@fkie.fraunhofer.de,nikita.jarocky@fkie.fraunhofer.de,frank.kurth@fkie.fraunhofer.de,sebastian.urrigshardt@fkie.fraunhofer.de,kevin.wilkinghoff@fkie.fraunhofer.de

doi:10.18420/inf2022_06

crime by identifying members and the connections between them to gather insights on the organization's structure as done in the *ROXANNE* project [RO]. To robustly recognize speakers, other components such as language, accent, gender and age identification can be utilized as done in the *Speaker Identification Integrated Project (SIIP)* [Kh18]. In both projects, field tests focusing on a user-centered design have been conducted with several law enforcement agencies through INTERPOL. Speaker diarization systems can also be used for post-processing police interviews [AF12]. Another example is the usage of surveillance systems in public places [Cr16]. These systems usually focus on acoustic events such as gun shots, breaking glass and screams and also try to localize sources.
Since annotating audio data is important for obtaining reliable training data in many applications, appropriate tools for such a task have been developed. Commercial tools, for instance, are provided by *appen* [Ap] or *Shaip* [Sh]. The appen audio annotation tool supports users by automatically segmenting audio files and allowing to add time stamps and transcriptions. Shaip offers a variety of label tools for text, image, video, and audio annotations. With SCALA-Speech, we propose a tool which allows for both annotating audio data as well as systematically analyzing its content by providing a variety of speech processing algorithms, all of which combined in a single speech analysis tool.
The contributions of this paper are the following: Several components needed for assisting an investigator in analyzing speech data are proposed and described. SCALA-Speech, a tool incorporating these components and realizing them with exemplary algorithms that can be selected depending on the application, is presented. As an example, using SCALA-Speech is demonstrated on audio extracted from videos contained in Twitter posts related to Belarus.

2 SCALA-Speech

2.1 Single Channel Analysis and Labeling Application (SCALA)

SCALA – a Single Channel Analysis and Labeling Application – is a tool developed with the purpose of analyzing and annotating different types of audio data. The main aspects include the application of integrated speech analysis algorithms and the visualization of their results. In addition, different kinds of annotated data may be visualized temporally aligned with a time-frequency representation (i.e., a spectrogram) of the audio signal itself. Visualizations include, in addition to displaying different types of annotations from label files, also feature matrices or special data types such as F0 trajectories. SCALA-Speech provides tools for signal analysis and enhancement, e.g. by applying filters or adjusting the carrier frequency as in the case of processing amplitude modulated RF (radio frequency) communication signals. A typical example of the graphical user interface (GUI) is shown in Fig. 1. The layout of the GUI can be chosen flexibly from several pre-defined layouts, depending on the given application. Individual GUI layouts may be created as well, allowing a user to reduce the view to the essential components required for the given scenario.
SCALA-Speech offers a large set of built-in speech processing algorithms, as specified in Sec. 2.2 and partially explained in more detail in the following sections. A short overview

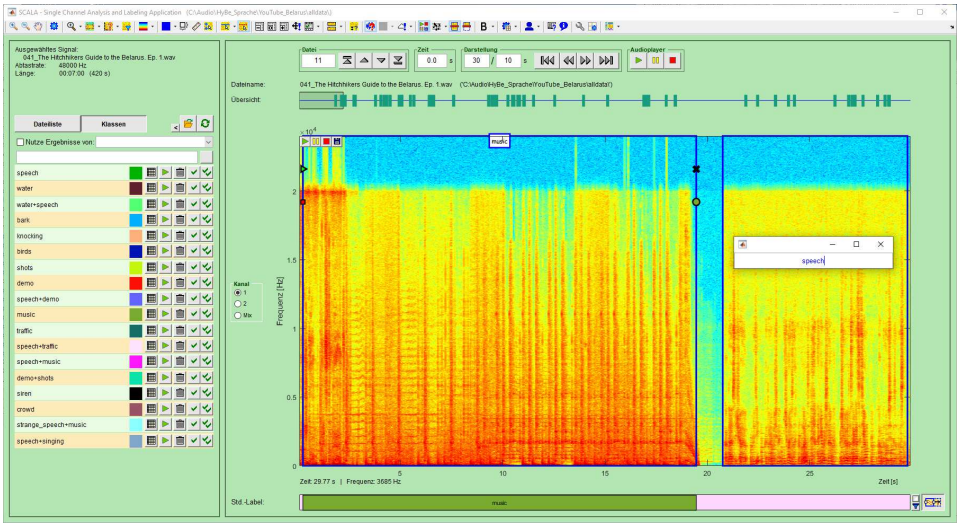


Fig. 1: SCALA-Speech: Layout in label mode generating a standard label file. The spectrogram contains two annotations, one in the process of being built, the left-hand side shows a class list.

about the covered speech analysis techniques is shown in Fig. 2 (left-hand side in blue). Next to the built-in methods, an excerpt of the basic tools is shown (right-hand side in green). Apart from analyzing signals by applying algorithms and visualizing results, one of the main tools provided is the *label mode*. Segments of signals may be marked by highlighting the corresponding regions of interest by simply specifying a rectangular region within the spectrogram. An example can be found in Fig. 1, depicting two rectangular regions in the spectrogram of the selected signal. The left-hand side shows a customized list of available classes. Annotations are stored in a user-defined format, e.g. in a standard label format including information about start time, end time, label, and – in case of modulated speech signals – carrier frequency, bandwidth, and sideband. In addition to machine-readable label files, the highlighted segments may also be saved as single wave files in class specific sub-folders thus allowing to train, for instance, keyword spotting or speaker recognition systems. For the latter case, the GUI offers a module for training speaker recognition models for several recognition methods. An additional feature allows for importing results of any algorithm applied to a signal as a starting point for manual annotations. Speech detection and segmentation, for instance, may provide relevant speech regions which can be labeled manually to generate speaker ground truth data. SCALA-Speech also includes a program for parameter optimization allowing to find optimal thresholds and parameter setups with respect to detection and false alarm rates. Finally, additional performance tracks may be shown.

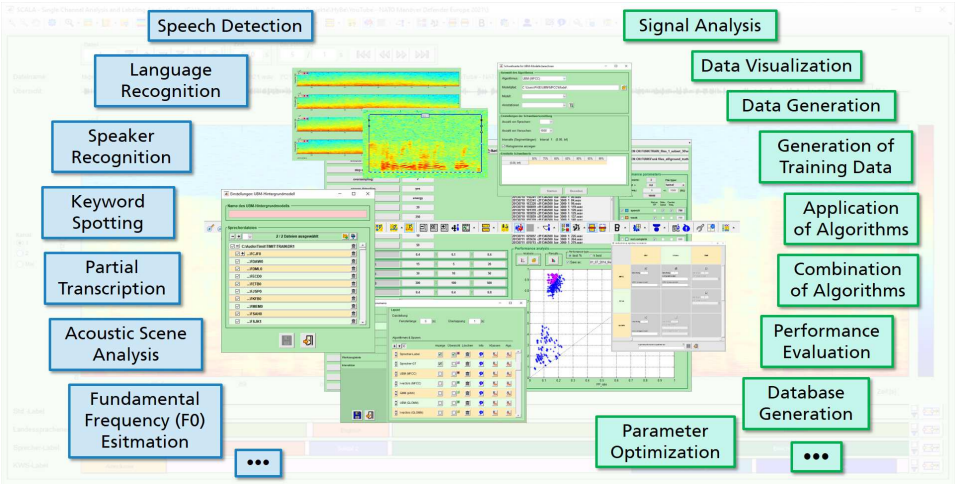


Fig. 2: SCALA-Speech: Overview about built-in components and tools. The system is modular and individual algorithms realizing these components can be selected depending on the application.

2.2 Speech Analysis Components

SCALA-Speech provides a user with a variety of audio processing modules, each of which realized by one or more algorithms particularly suited for specific speech analysis tasks. For e.g. speaker recognition, several different algorithms are available. An overview is given in Fig. 2 (left-hand side), including speech detection, audio segmentation, language recognition, speaker recognition, keyword spotting, as well as any combination of these. Note that these exemplary algorithms can be replaced with other algorithms depending on the particular application. Several included algorithms are described next.

2.2.1 Speech Detection

Speech detection (SD) or voice activity detection (VAD) is a mandatory pre-processing step for many speech analysis algorithms. In particular for heterogeneous data, it is important to extract those regions of audio signals containing speech, as opposed to other acoustic classes such as music, nature, or traffic noise. Depending on the given scenario, different SD algorithms may be used, each having different advantages depending on the given data. The so-called *Shift-F0* speech detection method extracts the fundamental frequency (F0) of speech by using an enhanced autocorrelation function (ACF) in the spectral domain [KCUU14, KCU14]. In contrast to the standard autocorrelation function, the shift-ACF applies a variable combination of product and minimum operations to extract multiply repeated components, while suppressing non-repeating structures [Ku13]. It is applied to the time dimension of a spectrogram, where the F0 appears as a horizontal energy-rich

component with additional energy at the respective harmonic frequencies. Those harmonics, being multiples of F0 (repeated structures in the spectral domain) can hence be detected by means of ACF-based methods. This approach is suited for RF communication scenarios that typically contain different and potentially strong noise types [KCU14, KCUU14].
VAD-MFCC is a voice activity detection (VAD) method based on Mel frequency cepstral coefficients (MFCC) in combination with spectral energy envelope extraction [vZ12]. This approach is a well-suited pre-processing step for further speech analysis algorithms such as speaker diarization or language recognition.It does not need any training material and is independent of speakers or language as opposed to statistical or learning-based approaches.
BBSD is a broadband speech detection algorithm developed for RF communication [UKK16]. It is composed of a combination of different methods that extract statistics of audio features such as spectral flux or spectral flatness. BBSD detects speech regions regardless of the given carrier frequency of the transmitted signals, distinguishing them from other RF signals.

2.2.2 Language Recognition

After detecting regions that contain speech, the next step is to identify *which* language is being spoken. For this purpose, we utilize a Speechbrain [Ra21] model based on the ESCAPA-TDNN achitecture [DTD20] to extract embeddings suitable for discriminating among languages. More concretely, the model is pre-trained on the Voxlingua107 dataset [VA21], which contains 6628 hours of speech from Youtube videos in 107 different languages. For specific use cases e.g. with fewer languages, another classifier can be trained on top by using the pre-trained embeddings belonging to only these languages as an input. Moreover, embeddings can be extracted for music datasets such as Orchset [BMG16] and environmental sounds such as the DCASE 2019 acoustic scene classification dataset [MHV19] and utilized as additional classes to obtain more robust representations in case the speech detection algorithms yield some false positive results.

2.2.3 Speaker Diarization and Recognition

Another important component of SCALA-Speech is the possibility to search for specific speakers (speaker recognition) or detect changes of speakers (speaker diarization [Pa22]). As in language recognition, the state-of-the-art in speaker recognition and diarization is to utilize discriminative embeddings. We use a pre-trained x-vector model [Sn18] from Speechbrain [Ra21] trained on VoxCeleb2 [CNZ18], which consists of over 1 million utterances from Youtube videos of approximately 6000 celebrities. It is also possible to train a model for extracting embeddings from scratch but this requires a reasonable amount of labeled data. For speaker diarization, the procedure presented in [Ar20] based on spectral clustering [NJW01] is used. More classical methods such as Universal Background Model (UBM) [RQD00] and i-vector [De11] are also available. For all methods, multiple complementary speech features can be utilized and their results can be combined [Wi18].

2.2.4 Keyword Spotting

HFCC-KWS is a keyword spotting (KWS) system based on MFCC-like audio features, the so-called *HFCC-ENS* features (human factor cepstral coefficients using energy normalized statistics) [KvZ10, vZKM10, ZU13]. It is a query-by-example approach using dynamic time warping (DTW) for matching. This method is independent of the spoken language or of the speakers, though the quality of the detection results depends much on the similarity of query and data. The best results are obtained when using queries of similar acoustic characteristics as the target signal. As these are not always available, this is a trade-off between performance and the amount of required training data.
A variation of the above method uses a pre-trained wav2vec 2.0 model [Ba20] for cross-lingual representation learning [Co21] from Speechbrain [Ra21] to extract representations as feature vectors. The actual KWS method is the same as the one used with HFCC-KWS.
DTW-KENS-KWS combines traditional methods with deep learning approaches by using discriminative *embeddings* extracted with a neural network as features and DTW for extracting time positions of matches [WCUG21]. Instead of extracting paths, DTW is only used to calculate score curves. The resulting score matrix is then post-processed and enhanced. Even though deep learning approaches typically require large amounts of training data, this method requires only a few examples of each keyword and thus can be considered a low-resource approach. The performance of the proposed system has been efficiently demonstrated on a dataset containing spoken coordinates. While HFCC-KWS is better suited for longer phrases (cf. [KvZ10]), this method can also deal with short queries.

3 Use case

One possible application of SCALA-Speech is speech monitoring in a big data context, i.e., in application scenarios where large amounts of data have to be analyzed and thus automatic procedures for supporting a user are desirable. Some of the algorithms described above have been combined and integrated in a system to detect *hybrid threats* [GSS21]. Hybrid threats or attacks in general denote the coordinated application of a mixture of measures from different domains in order to exploit vulnerabilities of an opponent. In [GSS21], hybrid threats are considered involving measures in the social media and cyber domains as well as potential attacks in the electromagnetic spectrum (EMS). The authors describe a system consisting of a variety of different sensors including both so-called local sensors such as an RF-scanner (radio frequency) or a passive radar, as well as virtual sensors for, e.g. cyber threat analysis or social media analysis, all of which gathering the retrieved information in a shared visualization tool. The speech analysis algorithms proposed in this paper are included into that approach by systematically analyzing the audio content retrieved from videos or blogs linked in twitter messages, while the textual parts of the messages are scanned by a social media analysis sensor. It is worth to mention that the latter sensor can also be used for detecting *fake news* [PS21]. Integrated in such a system, a speech analysis tool can provide additional information to the purely text-based social media analysis.

Regarding further applications of the toolset made available in SCALA-Speech, for instance keyword spotting or speaker identification can be used in contexts where little training or reference data is available. This can be the case when analyzing particular languages or dialects, where no or only very few training data is given, e.g. in extremely non-standard acoustic conditions with particular or strong background noise.
The proposed speech analysis tool allows a user to gather information of different types and analyze audio data in different use cases. In use cases requiring training data, e.g., when using AI-based algorithms, SCALA-Speech provides a possibility to generate such material. One example is one of the built-in speaker identification methods. The algorithm is trained on a large number of different speakers and needs to be adapted when looking for particular speakers. In such a case, a small amount of reference audio data is needed, which can be created with the corresponding labeling tool. The same holds for KWS scenarios.

3.1 Dataset

Several of the built-in speech analysis algorithms of SCALA-Speech have been applied to a particular dataset developed and annotated in the context of the hybrid threat analysis tool proposed in [GSS21]. It consists of a variety of different audio files extracted from video files referenced in twitter messages. In addition to personal blogs, these include mainly YouTube videos, all related to the topic *Belarus*. Since the considered files are linked to the given topic only by means of predefined keywords, in this case *Belarus*, the data to be analyzed may comprise any topic, also, for instance, videos related to the Eurovision song contest. Other files may simply contain music, e.g., if someone has linked a music video in a twitter post. Heterogeneity of the considered data implies particularly parameterized analysis methods. In a first step, the relevant data needs to be extracted by means of SD and VAD methods to separate speech content from the rest. A language recognition algorithm provides background information on the keywords and speakers to search for. The dataset contains 139 audio files with a total length of 39 hours, 92 of which contain speech. The files vary from 1 minute to 6.7 hours of duration. This dataset is only used for demonstration purposes and a systematic evaluation cannot be presented since the work is still in progress.

3.2 Demonstration

Fig. 3 shows a typical use case-specific GUI layout. In addition to the file list and the spectrogram of the selected signal, this layout includes data tracks for speech detection, language recognition, speaker diarization, and keyword spotting, which are all accompanied by a separate label track providing manual annotations. SCALA-Speech allows a user to navigate through a given set of files.To get an overview about the relevant information, the list of files can be filtered according to available results of a selected algorithm, e.g. speech detection. The GUI can be configured in order to allow a performance comparison of different speech processing methods applied to the same task. Fig. 3 depicts an example

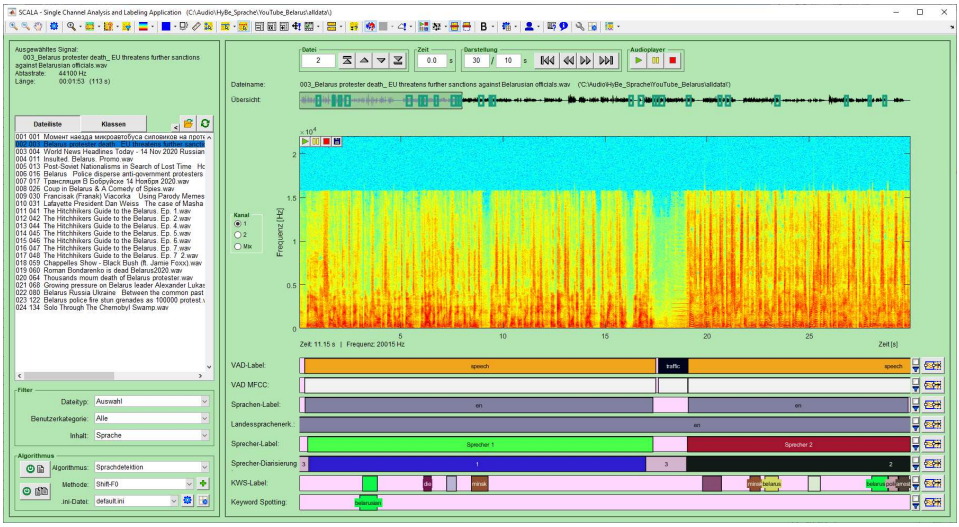


Fig. 3: SCALA-Speech algorithms: synchronous visualization of annotations and algorithm results.

taken from the described dataset. As can be seen, with the given data, most of the available speech analysis methods need to be adjusted, e.g. by additional training like in the case of keyword spotting or by parameter optimization as in the case of speech detection, and the language recognition approach needs to be combined with a VAD method. SCALA-Speech provides an overview on the quality of the data to be analyzed and represents an annotation tool for generating the necessary training or reference data.

4 Conclusions and Future Work

In this work, SCALA-Speech, a tool for assisting investigators in annotating and analyzing audio signals, has been presented. SCALA-Speech offers various modules for speech signal analysis such as voice activity detection, language recognition, speaker diarization/recognition and keyword spotting. The usage of SCALA-Speech has been demonstrated on audio tracks of videos obtained from Twitter posts related to Belarus.
For future work, systematic experimental evaluations with a focus on user experience for specific use cases are planned. To this end, we collaborate with partners representing different application domains. Some of the components have already been discussed with users in the domain of security-related applications. Based on the obtained results, individual signal analysis modules of SCALA-Speech will be adapted in order to meet the requirements of the particular use cases. The system will furthermore serve as a framework for integrating and testing audio and speech signal analysis algorithms developed in the future. Concerning mass data analysis, the audio analysis algorithms evaluated in SCALA-Speech can subsequently

be integrated in automatic signal production systems such as those we are developing in parallel in the context of RF signal analysis.

Bibliography

[AF12] Alexander, Anil; Forth, Oscar: Blind Speaker Clustering Using Phonetic and Spectral Features in Simulated and Realistic Police Interviews. In: IAFPA. 2012.

[Ap] Appen, `https://appen.com/data-types/#data-audio`, visited on 2022-05-12.

[Ar20] Aronowitz, Hagai et al.: New Advances in Speaker Diarization. In: INTERSPEECH. ISCA, pp. 279–283, 2020.

[Ba20] Baevski, Alexei et al.: wav2vec 2.0: A Framework for Self-Supervised Learning of Speech Representations. In: NeurIPS. 2020.

[BMG16] Bosch, Juan J.; Marxer, Ricard; Gómez, Emilia: Evaluation and combination of pitch estimation methods for melody extraction in symphonic classical music. Journal of New Music Research, 45(2):101–117, 2016.

[CNZ18] Chung, Joon Son; Nagrani, Arsha; Zisserman, Andrew: VoxCeleb2: Deep Speaker Recognition. In: INTERSPEECH. ISCA, pp. 1086–1090, 2018.

[Co21] Conneau, Alexis et al.: Unsupervised Cross-Lingual Representation Learning for Speech Recognition. In: INTERSPEECH. ISCA, pp. 2426–2430, 2021.

[Cr16] Crocco, Marco et al.: Audio surveillance: A systematic review. CSUR, 48(4):1–46, 2016.

[De11] Dehak, Najim et al.: Front-End Factor Analysis for Speaker Verification. IEEE Trans. Speech Audio Process., 19(4):788–798, 2011.

[Di21] Dikici, Erinç et al.: ROXSD: a Simulated Dataset of Communication in Organized Crime. In: SPSC. ISCA, pp. 32–36, 2021.

[DTD20] Desplanques, Brecht; Thienpondt, Jenthe; Demuynck, Kris: ECAPA-TDNN: Emphasized Channel Attention, Propagation and Aggregation in TDNN Based Speaker Verification. In: INTERSPEECH. ISCA, pp. 3830–3834, 2020.

[GSS21] Gerz, Michael; Schwarze, Arne; Stuch, Hans Peter: Connecting the Dots–Enhancing the Information Processing Chain for the Detection of Hybrid Threats for Host Nation Support and Territorial Operations. In: STO, NATO-OTAN. 2021.

[KCU14] Kurth, Frank; Cornaggia-Urrigshardt, Alessia: Detection of audio events with repetitive structure using generalized autocorrelations. In: ITG Speech. VDE, 2014.

[KCUU14] Kurth, Frank; Cornaggia-Urrigshardt, Alessia; Urrigshardt, Sebastian: Robust F0 estimation in noisy speech signals using shift autocorrelation. In: ICASSP. IEEE, pp. 1468–1472, 2014.

[Kh18] Khelif, Khaled et al.: SIIP: An innovative speaker identification approach for law enforcement agencies. In: STO, NATO-OTAN. pp. 1–14, 2018.

[Ku13] Kurth, Frank: The shift-ACF: Detecting multiply repeated signal components. In: WASPAA. IEEE, pp. 329–332, 2013.

[KvZ10] Kurth, Frank; von Zeddelmann, Dirk: An Analysis of MFCC-like Parametric Audio Features for Keyphrase Spotting Applications. In: ITG Speech. VDE, 2010.

[MHV19] Mesaros, Annamaria; Heittola, Toni; Virtanen, Tuomas: Acoustic Scene Classification in DCASE 2019 Challenge: Closed and Open Set Classification and Data Mismatch Setups. In: DCASE. pp. 164–168, 2019.

[NJW01] Ng, Andrew Y.; Jordan, Michael I.; Weiss, Yair: On Spectral Clustering: Analysis and an algorithm. In: NIPS]. MIT Press, pp. 849–856, 2001.

[Pa22] Park, Tae Jin et al.: A review of speaker diarization: Recent advances with deep learning. Comput. Speech Lang., 72:101317, 2022.

[PS21] Pritzkau, Albert; Schade, Ulrich: Vorsicht: mögliche „Fake News "–ein technischer Ansatz zur frühen Erkennung. In: Wahrheit und Fake im postfaktisch-digitalen Zeitalter, pp. 115–131. Springer, 2021.

[Ra21] Ravanelli, Mirco et al.: SpeechBrain: A General-Purpose Speech Toolkit. CoRR, abs/2106.04624, 2021.

[RO] ROXANNE Real time netwOrk, teXt and speaker ANalytics for combating orgaNized crimE, https://roxanne-euproject.org/, visited on 2022-05-16.

[RQD00] Reynolds, Douglas A.; Quatieri, Thomas F.; Dunn, Robert B.: Speaker Verification Using Adapted Gaussian Mixture Models. Digit. Signal Process., 10(1-3):19–41, 2000.

[Sh] Shaip, https://www.shaip.com/offerings/audio-annotation, visited on 2022-05-12.

[Sn18] Snyder, David et al.: Spoken Language Recognition using X-vectors. In: Odyssey. ISCA, pp. 105–111, 2018.

[UKK16] Urrigshardt, Sebastian; Kreuzer, Sebastian; Kurth, Frank: General detection of speech signals in the time-frequency plane. In: ITG Speech. VDE, 2016.

[VA21] Valk, Jörgen; Alumäe, Tanel: VOXLINGUA107: A Dataset for Spoken Language Recognition. In: SLT. IEEE, pp. 652–658, 2021.

[vZ12] von Zeddelmann, Dirk: A feature-based approach to noise robust speech detection. In: ITG Speech. VDE, 2012.

[vZKM10] von Zeddelmann, Dirk; Kurth, Frank; Müller, Meinard: Vergleich von Matching-Techniken für die Detektion gesprochener Phrasen. DAGA, pp. 257–260, 2010.

[WCUG21] Wilkinghoff, Kevin; Cornaggia-Urrigshardt, Alessia; Gökgöz, Fahrettin: Two-Dimensional Embeddings for Low-Resource Keyword Spotting Based on Dynamic Time Warping. In: ITG Speech. VDE, pp. 9–13, 2021.

[Wi18] Wilkinghoff, Kevin; Baggenstoss, Paul M.; Cornaggia-Urrigshardt, Alessia; Kurth, Frank: Robust Speaker Identification by Fusing Classification Scores with a Neural Network. In: ITG Speech. VDE, pp. 261–265, 2018.

[ZU13] Zeddelmann, Dirk von; Urrigshardt, Sebastian: Ein Demonstrator zum Keyword-Spotting basierend auf gehörangepassten Audiomerkmalen. In: GI-Jahrestagung. Gesellschaft für Informatik e.V., pp. 3026–3028, 2013.

A frequentist estimation of duplicate probability as a baseline for person identification from image and video material using anthropometric measurements

Florian Heinke[1] Marie-Luise Heuschkel[1] Dirk Labudde[1]

Abstract: Video and image material is becoming increasingly ubiquitous thus its potential as evidence in forensic investigations is growing. Once faces are hidden however, the value of surveillance footage is restricted unless there is another biometric trait that can be observed by camera such as linear body measurements. There is much biological evidence for human body proportions exhibiting much individual variation. Nevertheless, the probability of there being two individuals that match in their respective proportions ultimately determines its usability for the assignment, exclusion and even identification of persons in the forensic domain. This work is concerned with approaches for duplicate probability estimations derived from anthropometric measures.
Here, we present a novel frequentist estimation using a dataset of 340 individuals and their respective anthropometric measurements. Drawing on density kernel estimations of measurement dissimilarity, we propose the duplicate probability to be in the order of 10^{-15} to 10^{-8}.

Keywords: Biometric features; anthropometric measurements; duplicate probability

1 Introduction and Motivation

Biometric research increasingly goes beyond established biometric traits such as fingerprints, DNA, iris, and facial characteristics [RSV19] investigating the body for other traits such as linear body measurements i.e. heights, lengths and widths [ONS10] that could be used for matching and identification in order to increase and diversify possible use cases [Da11]. The idea is not new as anthropometric measurements had been part of Bertillonage, a pioneer biometric system developed at the end of the 19th century [Al18]. A reason why this system did not last and was surpassed by fingerprinting some 30 years later, was that such measurements - in contrast to fingerprints - were not left as traces at a crime scene [Co07, DC08, Ka17]. With the ever-increasing amount of image and video material available that is no longer true since a person's body shape including their dimensions and proportions can now be caught on camera. Indeed, when the face is concealed as it often happens during a crime, they might be the only hint at a person's identity [AA14]. Body dimensions and proportions are universal with the only limitation being the number of features, for instance brought about by the loss of a limb. Even though temporary and long-term fluctuation, e.g. resulting in a slight decrease in body height over the course of the

[1] University of Applied Sciences Mittweida, Department of Natural and Computer Sciences, Technikumplatz 17, 09648 Mittweida, Germany heinke@hs-mittweida.de

doi:10.18420/inf2022_07

day and in old age respectively [LTZ87], must be accounted for, linear body dimensions that equal bone lengths can be considered quite permanent [SS13] as soon as an individual is fully grown between 18 and 21 years of age [SSS14, LTZ87]. The literature provides ample evidence that body measurements and their respective proportions show much individual variation [He60, JMW98, Gr12, SS13] ensuing in people of the same height having different proportions [Lu98, BHM16] or all kinds of proportions occurring across all heights [Ka69]. These differences become apparent even within a homogenous population [Bu15]. However, as this kind of evidence remains exemplary in nature, eventually it is up to the probability of there being two individuals that match in their respective proportions that determines whether linear body measurements may qualify as a biometric identifier. With only limited information insufficient for statistic model building available the challenge is to ensure a correct probability estimation which will be the focus of this paper.

2 Datasets and Methods

2.1 Considered Data

The data is a subset of the SizeGermany, a nationwide survey that was conducted between 2007 and 2009 and includes more than 13,000 3D bodyscans (average error less than 1 mm [circumference]) as well as anthropometric measurements that were derived from these scans [RT10, STW09]. Provided by Avalution GmbH, the subset consists of eleven height, length and width measurements compliant with the ISO 8559, ISO 7250 and RAMSIS (associated with the RAMSIS digital human model) standards [RT10] and of high-resolution 3D bodyscans taken from 170 female and 170 male individuals. The data was further selected with regard to age (18-55 years) and two body height classes for each of the sexes. Thus, it is comprised of four groups (labelled F1, F2, M1 and M2 in the following sections) here group F1 corresponds to 85 females with a measured body height of approx. 163 cm (sd = 0.13 cm, max. deviation: 0.2 cm). F2 consists of 85 females at body height 173 cm (sd = 0.14 cm, max. dev.: 0.2 cm). For male individuals, measurements of 85 men with an approx. body height of 175 cm (group M1, sd = 0.17 cm, max. dev.: 0.3 cm) respectively 185 cm (group M2, sd = 0.22 cm, max. dev.: 0.4 cm) were acquired. The grouping of four subsets was intended for exploratory analysis of intra- and inter-group correlations with respect to the eleven measurement features.
Table 1 gives an overview of the features considered in this study as well as the minimum measurement resolutions.

2.2 Probability estimations

One aim of this study is to give an estimate on the probability for obtaining *matching* measurement values for two unrelated individuals with respect to all considered features,

Tab. 1: Anthropometric measurement features, applied measurement standards and minimum resolution per feature.

Feature description	Measurement std.	Resolution (cm)
Stature (body height)	ISO 7250 Nr. 4.1.02	0.1000
Shoulder height (standing)	ISO 7250 Nr. 4.1.04	0.0500
Tibial height	ISO 7250 Nr. 4.1.08	0.1000
Shoulder-elbow length	ISO 7250 Nr. 4.2.06	0.0125
Elbow-wrist length	ISO 7250 Nr. 4.2.07	0.0500
Shoulder (biacromical) breadth	ISO 7250 Nr. 4.2.08	0.1000
Forearm-fingertip length	ISO 7250 Nr. 4.4.05	0.1000
Buttock-knee length	ISO 7250 Nr. 4.4.07	0.0500
Pelvis breadth	RAMSIS 1.1	0.1000
Waist Height	ISO 8559 Nr. 2.2.3	0.1000
Outside Leg Length	ISO 8559 Nr. 2.2.25	0.0500
Inside Leg Length	ISO 8559 Nr. 2.2.27	0.1000

based on the data at hand. First, one has to note the notion of what is actually meant by stating that two *exactly identical* measurements are obtained: Regarding to experimental data acquisition one has to account experimental errors and deviations as well as the measurement method's specific resolution itself, which leaves one with feature scale intervals and length thresholds within respectively below no meaningful statements about measurement deviations can be made. Secondly, deriving (and experimentally confirming) such a probability estimation based on first principles requires to understand all underlying stochastic and correlated processes, and to enclose that knowledge in a self-consistent statistical model, which in turn provides the means for making such estimations. These two aspects result to the basis at which the match estimations are made in this study: First, identity ought to be assumed if absolute measurement differences between two individuals are below the respective feature resolutions (and experimental measurements would trivially read the same values). On a more formal note, we here propose the euclidean distance $d(\mathbf{x}_i, \mathbf{x}_j)$, defined as

$$d(\mathbf{x}_i, \mathbf{x}_j) = \sqrt{\sum_{l=1}^{11} (x_{i,l} - x_{j,l})^2} \tag{1}$$

as a function for measurement dissimilarity, where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the eleven measurements of two individuals. With this definition we can reflect the first notion in a more formal manner: With respect to the feature resolutions given in Table 1, there is a minimum, non-zero distance d_{min} which can be experimentally measured and below which identity for two given $\mathbf{x}_i$ and $\mathbf{x}_j$ can be assumed. Regarding the second notion, this study focusses on providing match probability estimations straightforwardly from distance distributions, thus without to

consider underlying feature correlations or to state any statistical model.
Exploratory data analysis indicated that the distributions of intra-group measurement distances, as defined in Equation 1, can be described by means of the gamma distribution. Subfigure 1a shows the four distributions of the $N(N-1)/2$ pairwise measurement distances for each group F1, F2, M1 and M2, where $N = 85$ in each group. A straightforward approach to give an estimate on the probability for obtaining two identical measurements (in the sense as noted above) at random–denoted as p_{match} in the following text–is to first obtain the cumulative gamma density function parameters α and β which fit a given set of pairwise distances D best, and second to retrieve p_{match} from the fitted distribution by numerical integration. Thus, $p_{\text{match}} = F(d < d_{min}|\alpha, \beta)$, where F is the cumulative density function of the gamma distribution. However, the four initial distance distributions are noisy to some extend, as seen in Figure 1a. Hence following steps 1 and 2 on these four distributions would give four point estimates prone to error. A further source for error is that each of the four sets of pairwise distances D holds $d_{min} << D$, whereas the computed density estimation $F(d < d_{min}|\alpha, \beta)$ is many orders smaller than $2/N(N-1)$ and thus highly sensitive to the fitted distribution parameters. To cope with these problems, we conducted boostrapping with $n = 250$ runs on each distance set D. In each run, half of the pairwise distances were sampled randomly and steps 1 and 2 were conducted on this subset. This strategy resulted in 250 bootstrap distribution fits for each sampled pairwise distance set. The fitted distributions were used in turn for computing 250 p_{match} estimations. Furthermore, note that neither in nor between all four groups a match of two individuals was observed. In Figure 1b and c, 250 goodness-of-fit statistics (expressed by harmonic relative mean errors) and the resulting distributions of p_{match} estimations are shown.

3 Results and Discussion

Based on bootstrapping approximation, p_{match} estimates range from the order of 10^{-15} to 10^{-8} over all four investigated intra-group distances (1st quartile: 2.2×10^{-12}, median: 1.2×10^{-11}, 3rd quartile: 9.4×10^{-11}). The harmonic relative mean errors indicate that observed distribution frequencies are approximated well by all fitted density functions. In fact, all fitted functions give estimates deviating by less than 1% wrt to the harmonic relative mean error (see Figure 1b). As no matching pair of individuals within nor between the four groups was observed with respect to all eleven features, there are only limited ways to further test obtained probability estimations. A simple testing approach is to interpret the observation of obtaining $k = 0$ matches over the total number of pairwise comparisons N_{total} as the result of some binomial process. Following this assumption one can estimate the probability of obtaining at least one match by chance using a binomial distribution $P_{k\geq 1} \approx 1 - B(N_{\text{total}}, k = 0, p = p_{\text{match}})$, where $N_{\text{total}} = 340 \cdot (340-1)/2$. Plugging in above estimates into the binomial gives approximate probabilities in the order of $P_{k\geq 1} \approx 10^{-7}$ to 10^{-6}. Conversely, a conservative assumption that $P_{k\geq 1} = 0.5$ or 0.05 would correspond to baseline approximations of $\tilde{p}_{\text{match}} \approx 1.2 \times 10^{-5}$ and 5.2×10^{-5}, respectively. As the $P_{k\geq 1}$ estimations is consistent with observing $k = 0$ on the one hand, and the conservative

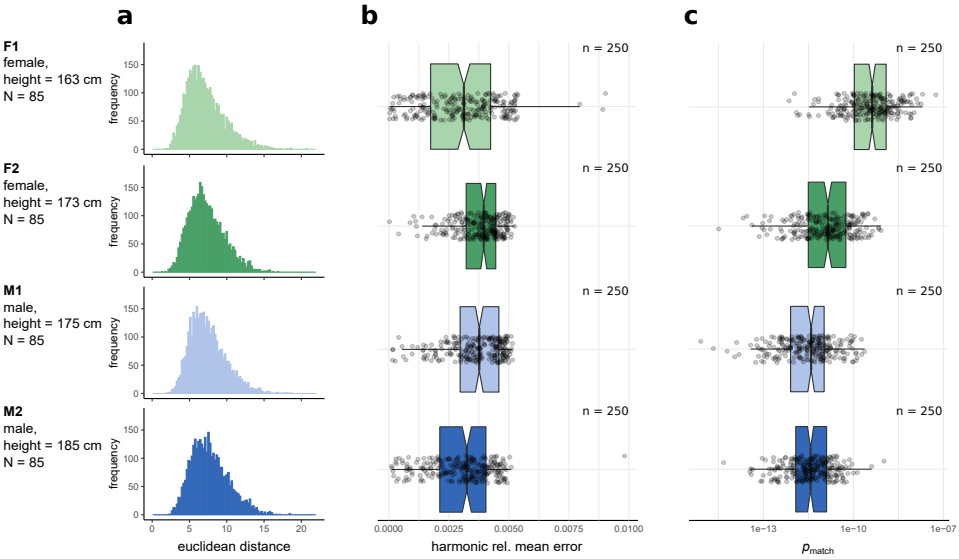


Fig. 1: Frequency distributions of intra-group feature dissimilarities (a), expressed by the euclidean distance, provide the basis for model-free duplicate probability estimation. b: Goodness-of-fit statistics (as harmonic relative mean error) of Gamma distributions fitted to 250 dissimilarity frequency distributions obtained from each group by bootstrapping. c: Duplicate probability estimations obtained from bootstrapping. Estimations range in the order of 10^{-15} to 10^{-8} over all four groups.

$\tilde{p}_{\text{match}}$ approximations being too large to be compatible with density estimations, one can cautiously regard the p_{match} estimate range as reasonable. Next, this consistency check was further applied to subsets of features for which identities can be observed, in order to verify if the proposed p_{match} estimation approach holds when $k > 0$. For example, such a set of features is {Tibial height, Elbow wrist length, Buttock knee length}. Considering only these three features, a single match can be observed between groups F2 and M1. Estimating p_{match} based on the inter-group distance distribution, one obtains $P_{k\geq 1} \approx 0.32$ (std. dev: 0.04) under the binomial model, which in turn is compatible with the observation. This gives an indication not just that match probability estimations might be reasonable, but also that the model-free estimation strategy presented here is compatible with the goal of making estimations in the first place. Further research will involve investigating the reasoning behind dissimilarities following the gamma distribution–which is some aspect the general notion of the gamma distribution is hardly conform with.
Putting the presented results into perspective, one first has to reflect that the proposed estimation strategy is tied to the empirically determined measurement distance distribution and the underlying measurement features inspected. However, there is indication that, given the data and features investigated, match probabilities might be in the order of magnitude where forensic applicability can be reasonable.

4 Conclusion and Outlook

In this work, an estimation for feature match probabilities is given based on experimental anthropometric measurement features and intra-group measurement dissimilarity distributions. These probability estimates give some insights into whether anthropometric measurements can be valuable in forensic applications, e.g. person identification. The results presented in this study are an indication of that.
Going hand in hand with the advances in the field of 3D digitalisation and modelling that allows for these proportions to be extracted from digital material, this estimation may serve as a foundation for further application of anthropometric measurements in the forensic domain, especially when it comes to the detection and measurement of individuals depicted in images and videos for the purpose of some level of identification thus rendering novel biometric methodologies feasible.
A critical aspect to consider is that, the here presented estimations are 'statements on the statistical mean' and that proper systematic estimations would require knowledge about the fundamental correlations and dynamics that drive anthropometric measure variations in individuals and, most importantly, determine a person's specific measurements. Such first principle model-based estimations are of most value in the forensic context, since statistical statements are not just made upon mean estimates deduced from empirical data, but are rather embedded in a theoretical statistical framework capable to include other forensic evidence as well.

Bibliography

[AA14] Andersson, V. O.; Araújo, R. M.: Full Body Person Identification Using the Kinect Sensor. In: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence, Lymassol, Cyprus. pp. 627–633, 2014.

[Al18] Altes, K. B.; Ost, A.; Perez, D.; Constantino, A.; Carpenter, K.; Ferguson, D.; Bohne, C.: Bertillon, Alphonse. In (Trevathan, W.; Cartmill, M.; Dufour, D.; Larsen, C.; O'Rourke, D.; Rosenberg, K.; Strier, K., eds): The International Encyclopedia of Biological Anthropology. John Wiley & Sons, Inc, Hoboken, NJ, USA, 2018.

[BHM16] Bullinger-Hoffmann, A. C.; Mühlstedt, J.: Homo Sapiens Digitalis - Virtuelle Ergonomie und digitale Menschmodelle. Springer Berlin Heidelberg, 2016.

[Bu15] Bubb, H.; Bengler, K.; Grünen, R. E.; Vollrath, M.: Automobilergonomie. Springer Fachmedien Wiesbaden, 2015.

[Co07] Cole, S. A.: Twins, Twain, Galton, and Gilman: Fingerprinting, Individualization, Brotherhood, and Race in Pudd'nhead Wilson. Configurations, 15(3):227–265, 2007.

[Da11] Dantcheva, A.; Velardo, C.; D'Angelo, A.; Dugelay, J.-L.: Bag of soft biometrics for person identification. Multimedia Tools and Applications, 51(2):739–777, 2011.

[DC08] Dessimoz, D.; Champod, C.: Linkages between Biometrics and Forensic Science. In (Jain, Anil K.; Flynn, Patrick; Ross, Arun A., eds): Handbook of biometrics. Springer International Publishing Cham, Switzerland, 2008.

[Gr12] Grupe, G.; Christiansen, K.; Schröder, I.; Wittwer-Backofen, U.: Anthropologie: Einführendes Lehrbuch. Springer Berlin Heidelberg, 2012.

[He60] Hertzberg, H. T. E.: Dynamic Anthropometry of Working Positions. Human Factors: The Journal of the Human Factors and Ergonomics Society, 2(3):147–155, 1960.

[JMW98] Jürgens, H. W.; Matzdorff, I.; Windberg, J.: Internationale anthropometrische Daten als Voraussetzung für die Gestaltung von Arbeitsplätzen und Maschinen. Arbeitswissenschaftliche Erkenntnisse, 108, 1998.

[Ka69] Kadanoff, D.: Körperbau und Konstitutionstypen des Menschen. Anthropologischer Anzeiger, 31(3):193–202, 1969.

[Ka17] Kanwar, A. R.: Anthropometry and its Historical Development. International Journal of Researches in Social Sciences and Information Studies, V(March):33–37, 2017.

[LTZ87] Leonhardt, H.; Tillmann, B.and Töndury, G.; Zilles, K., eds. Rauber/Kopsch: Anatomie des Menschen: Lehrbuch und Atlas. Thieme Stuttgart, 1987.

[Lu98] Luczak, H.: Arbeitswissenschaft. Springer Berlin Heidelberg, 1998.

[ONS10] Ober, D. B.; Neugebauer, S. P.; Sallee, P. A.: Training and feature-reduction techniques for human identification using anthropometry. In: 2010 4th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) Washington, DC, USA. IEEE, pp. 1–8, 2010.

[RSV19] Ramalingam, S.; Shenoy, A.; Viet, N. T.: Fundamentals and Advances in 3D Face Recognition. In (Obaidat, M. S.; Traore, I.; Woungang, I., eds): Biometric-Based Physical and Cybersecurity Systems, pp. 125–162. Springer International Publishing Cham, Switzerland, 2019.

[RT10] Rissiek, A.; Trieb, R.: iSize - Implementation of International Anthropometric Survey Results for Worldwide Sizing and Fit Optimization in the Apparel Industry. In (D'Apuzzo, Nicola, ed.): Proceedings of the International Conference on 3D Body Scanning Technologies Lugano, Switzerland, 19-20 October 2010. Hometrica Consulting N. D'Apuzzo, pp. 269–281, 2010.

[SS13] Scheffler, C.; Schüler, G.: KAN-Studie 51 Rohfassung eines Leitfadens für die richtige Auswahl und Anwendung anthropometrischer Daten. KAN Kommission Arbeitsschutz und Normung Sankt Augustin, 2013.

[SSS14] Schünke, M.; Schulte, E.; Schumacher, U.: Prometheus: LernAtlas der Anatomie: Allgemeine Anatomie und Bewegungssystem. Prometheus. Thieme Stuttgart, 4th edition, 2014.

[STW09] Seidl, A.; Trieb, R.; Wirsching, H.-J.: SizeGERMANY–the New German Anthropometric Survey conceptual design, implementation and results. In: Proceedings of 17th World Congress on Ergonomics. 2009.

Challenges of Network Traffic Classification Using Deep Learning in Virtual Networks

Daniel Spiekermann[1], Jörg Keller[2]

Abstract: The increasing number of network-based attacks like denial-of-service and ransomware have become a serious threat in nowadays digital infrastructures. Therefore, the monitoring of network communications and the classification of network packets is a critical process when protecting the environment. Modern techniques like deep learning aim to help the providers when detecting anomalies or attacks by learning details extracted from a network packet or a flow of packets. Most of these models are trained in networks without any kind of virtualisation, especially network virtualisation overlay environments are not investigated in detail. In this paper, we analyse the classification rate of a Convolutional Neural Network (CNN) faced with encapsulated packets. We evaluate this approach with a proof-of-concept based on a binary classification of a self-curated data-set.

Keywords: virtual networks, network virtualisation overlay, deep learning, neural networks, network traffic classification

1 Introduction

The use of network virtualisation overlay (NVO) is a common technique in modern network infrastructures. A static network design does not provide the required flexibility for modern environments, which demand for dynamic and ubiquitous network access, high speed packet transfers, east-west traffic, customizability and secure tenant isolation [BAM10]. With the evolution of dynamic overlay protocols, limitations of traditional networks were eradicated and the flexibility inside the network as well as the automation of the environment increases. Overlay protocols like VXLAN, GENEVE or NVGRE encapsulate the original network packet by preceding one or more additional header information. Overlay protocols increase the possibilities inside the network, but add complexity for involved devices like switches, routers and firewalls when analyzing network packets because of the additional layers. In addition to this, security staff, digital forensic investigators or support teams are faced with a higher complexity when capturing and analysing the transferred network protocols, i. e. for troubleshooting, malware detection or law enforcement.

Typically, a provider uses only one of these protocols in the environment, but the use of different protocols simultaneously or the combination of protocols at the same time is possible. Using one or more of these encapsulating protocols results in changed network

[1] Polizeiakademie Niedersachsen, Germany daniel.spiekermann@polizei.niedersachsen.de
[2] FernUniversität in Hagen, Germany joerg.keller@fernuni-hagen.de

doi:10.18420/inf2022_08

packets inside the environment. Whereas network packets inside specific network areas (e. g. a tenant network) remain without any overlay protocol, packets transferred inside the backbone network might be changed depending on the path they are traversing. Figure 1 shows an exemplary network with only two different tenants and three network segments.

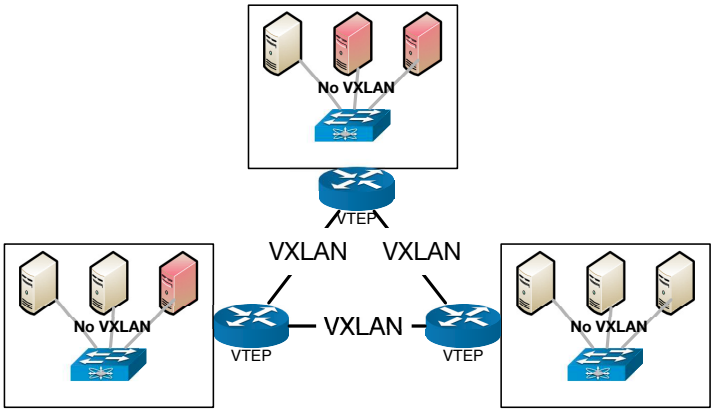


Fig. 1: NVO usage in networks, machines of different tenants are marked in different colours. Each rectancle represents a single subnet (e. g. a rack or compute node), the VTEPs manage the communication of nodes of the same color.

Inside a tenant network, e. g. the upper square with two red VMs, no VXLAN encapsulation is needed, but if the red VM in the left square is connected, each network packet is encapsulated if sent to this VM. Depending on the position of capturing, this protocol change effects all types of network classification, intrusion detection or security devices like firewalls and application layer gateways.

Providers of modern networks use traffic classification techniques to get insights about communication inside the environment. Static approaches like deep-packet inspection as well as dynamic techniques like machine learning or deep learning are used to increase the security of the infrastructure. Anomalies, outliers, adversarial attacks and unwanted traffic can be classified for further processing. So, the knowledge of network traffic inside an environment is crucial for every modern infrastructure.

Prior implementations mostly focus on a static ruleset to classify benign or malicious network traffic. As shown in [SK20], a shift from a static to a dynamic network infrastructure might result in more complex techniques, so static approaches are now insufficient.

With the evolution of machine learning approaches, the possibilities of network traffic classification, anomaly detection and the overall network-security increases. Different algorithms provide improved techniques to eradicate common issues like the heterogeneous traffic or user-customized network infrastructures. But these techniques still have issues

when faced with virtualised network traffic, which demands for new or adapted training sets to avoid inappropriate detection and protection results. As an example, [SK21b] details, that different algorithms like IsolationForest have fail in the detection of anomalies when faced with virtualised traffic. Using deep learning algorithms further enlarge the possibilities of artificial intelligence. But similar to machine learning algorithms, most of these applications do not cover the aforementioned virtual environments.

Deep learning techniques use multiple layers in an artificial network to provide various applications like image classification, speech recognition, recommendation systems or fraud detection, which gain an advantage over human or static implementations [GBC16]. In network environments, techniques like Convolutional Neural Networks (CNN) or Generative Adversarial Network (GAN) are used to improve the security [SSE19] of intrusion detection systems [Ta16], detect malware [Vi19] or spam [He17] and classify the network traffic [XLJ21].

An improved detection rate in network classification is reported in the literature [Lo20], but most of the research is focused on plain or encrypted network traffic. In this paper, we analyze the application of a CNN when classifying NVO traffic. As done in [Li19] for networks without virtualization, we convert network packets to pictures and train our CNN with these information. The detection rate of $\geq 99\%$ is similar to approaches like [Li19, Wa18, MMS19]. However, the analysis of the same network traffic encapsulated with different NVO protocols shows a reduced detection rate depending on the protocol.

The main contribution of our research are:

- We identify the role of different network virtualization protocols on anomaly detection algorithms.
- We explore the difficulties that deep learning algorithms encounter when being applied on virtual instead of physical networks.
- We provide a proof-of-concept to support our hypotheses with quantitative experiments. We provide a self-curated data-set for these experiments which we plan to make public.

The remainder of this article is structured as follows. Section 2 discusses related work regarding network traffic classification and virtual networks. In Section 3 we describe the initial classification process with a CNN. The trained model is used in Section 4 to determine the detection rate with encapsulated network packets. Section 6 concludes this paper and gives an outlook to our future research.

2 Related Work

Network traffic classification is a well-known and deeply researched area in all fields of information security like anomaly detection [Ma03] and attack prevention [Fe03] as well as network management and monitoring [Ts18]. [AKO15] gives an overview of different traffic classification techniques. Static techniques like port-based, payload-based algorithms or deep-packet inspection are discussed in [Ac10].

The use of deep learning as discussed in [Li19]. [Wa18] shows the possibilities of a CNN and residual network (ResNet) to classify network traffic, even when the traffic is encrypted.

NVO changes the internal structure of network infrastructures, which might result in critical effects for the detection algorithms [SK20]. The impact of an unsupervised packet based approach is investigated in [SK21c].

3 Deep learning classification

The use of deep learning is a common technique to classify network traffic. Most of the research is done without any dynamic encapsulation protocol [SSJ21, Ac19] but on plain network protocols. In addition to this, deep learning is able to classify network traffic even when it is encrypted or non-plaintext [Ac18]. By this, anomalous network traffic and malware might be detected and discarded or the knowledge of transferred network packets helps to improve the performance in a network.

To evaluate the impact of a network change to a CNN, [Ze19] proposes the following steps: *Packet Generation, Traffic Purification, Traffic Refiner, Length Unification* and *Packet conversion*. To generate network packets, we created a virtual network environment, which facilitates network packet captures at various positions. Inside this network, we focus on the transmission of ICMP (Internet Control Message Protocol) packets. These packets differ from a huge number of common network packets like TCP or UDP traffic which simplifies the initial process of network packet classification. So, we were able to improve the analysis of the classification process. The traffic purification and refinement was done by sanitizing the network traffic to remove unwanted network traffic like ARP requests or other broadcast packets. After this, our capture files contain only wanted network packets. Our first capture file contains 14,072 ICMP messages with different types like ICMP request (type 0) and response (type 8) or destination unreachable (type 3) [Po81]. The next step was the transformation of every single packet to a picture representing this packet. We use an adaptation of the Python module file2image[3], which ensures the correct transformation of the packet to a picture in png-format. Every network packet was scaled to 128×128 pixels, a result is shown in Figure 2.

We use Tensorflow and Keras[4] to create a CNN for image classification. We use the following

[3] Details can be found at `https://pypi.org/project/file2image/`
[4] Details can be found at `https://keras.io/about/`

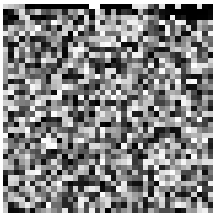


Fig. 2: Image of a network packet generated with file2image and scaled to 128×128 pixels

layers derived from [KJ18, SHA17] to create our model, which is trained with 50 epochs, but without any special optimisations: *Conv2D, MaxPooling2D, Activation, Dense, Add, Dropout* and *BatchNormalization*.

Our binary classification is used to detect ICMP packets in a number of different network packets and has a validation accuracy of 99%.

After this we analyse the model with a second data-set containing various network packets and protocols like HTTP, RTP, SSH and ICMP, but no NVO traffic. This data-set M comprises 2,761,441 network packets, the subset of ICMP packets is denoted by $I \subset M$. By enumerating each picture we are able to verify the prediction for each network protocol more easily.

Our model analyses each network protocol and predicts the classification rate, i.e. the probability p_i that packet i an ICMP packet. The following excerpt exemplifies the output of the analysis.

```
684.png is 99.97 percent NO ICMP and 0.03 percent ICMP.
685.png is 0.03 percent NO ICMP and 99.97 percent ICMP.
686.png is 99.99 percent NO ICMP and 0.01 percent ICMP.
687.png is 0.02 percent NO ICMP and 99.98 percent ICMP.
```

We use *tshark* to validate this results. Packets 685 and 688 are correctly classified as ICMP packets by the model, the other network packets are correctly classified as non-ICMP, because they are RTP, UDP and TCP packets. The following snipped illustrates how the predictions are validated.

```
tshark -r capture.pcap   -T fields -e frame.number -e_ws.col.Protocol
684 RTP
685 ICMP
686 UDP
687 ICMP
```

This demonstrates the correctness of our model, which has an average detection rate $\sum_{i \in I} p_i / |I|$ of 99.98%.

4 Packet transformation

This section describes the process of creating virtual network packets and the classification of these adapted packets with the already trained model.

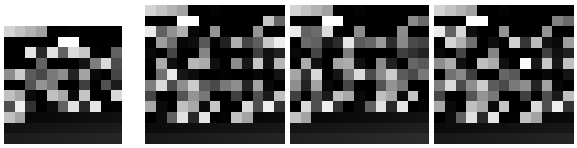
The shift of network packets to virtual packets is done with *Encapcap* [SK21a]. This tool helps to create encapsulated network packets from plain network packets by adding necessary header information and is able to fill randomized, but still plausible values into the new headers of these network packets, which simulates virtual behaviour like protocol swapping or changes of the IP addressing scheme. We transformed different network captures containing network packets with protocols (HTTP, HTTPS, ICMP, RTP, SSH and QUIC) with *Encapcap* to data-sets with encapsulating network protocols. We use VXLAN, NVGRE and GENEVE as NVO protocols.

- **VXLAN** The most notable protocol to implement NVO is virtual eXtensible LAN (VXLAN) [Ma14], which is similar to the well-known VLAN protocol, but expands its features and increases the number of possible virtual separated subnets to 16,777,216 networks. VXLAN uses a 24-bit header field named virtual network identifier (VNI) to isolate the different networks, the encapsulation is done with the User Datagram Protocol (UDP).
- **NVGRE** Network Virtualization using Generic Routing Encapsulation (NVGRE) is defined in RFC 7637 [GW15]. It bases on GRE as the encapsulating protocol and uses parts of the packet header to manage and control the network separation. NVGRE uses 24 bit of the optional fields to add a Virtual Subnet ID (VSID).
- **GENEVE** Generic Network Virtualization Encapsulation (GENEVE) is defined in [GGS20] and uses a 24-bit header field named Virtual Network Identifier (VNI). In addition to this, GENEVE provides huge flexibility for the encapsulation of different network packets. GENEVE uses a small fixed header followed by variable-length option fields used for the transmission of specific information like identifiers. Similar to VXLAN, GENEVE uses UDP as the transport protocol.

All implementations use a 24-bit header field for network isolation, but in combination with various header fields a resulting network packet differs after the encapsulation process depending on the used virtualization protocol.

We created one capture file for each aforementioned virtualization protocol. Each capture file contains the correct number of network packets. As described in Section 3, each network packet of each capture file is transformed to an image. Due to different protocols and their

type of encapsulation, different pictures result from this process. Figure 3 shows the different pictures.



(a) Plain (b) VXLAN (c) NVGRE (d) GENEVE

Fig. 3: Pictures of network packet before and after transformation

The transformed packets have a bigger size because the creation process takes care of the packet length and therefore creates pictures with a size depending on the packet size, and thus depending on the virtualization protocol. To ensure compatibility, we scale each image to 128×128 pixels.

5 Discussion

We use our trained model to classify each network packet i in the capture file M_P (the subset of ICMP packets is denoted as I_P) for virtualization protocol P either as ICMP or as non-ICMP, i.e. for each packet i we get probability p'_i of being an ICMP packet. Table 1 shows the average $\sum_{i \in I_P} p'_i / |I_P|$, best $\max_{i \in I_P} p'_i$ and worst $\min_{i \in I_P} p'_i$ detection rates for each protocol P. For comparison, we repeat that the average detection rate of plain network packets was 99,98%, as listed in Section 3.

Tab. 1: Detection rate

Protocol	**Average**	**Best**	**Worst**
VXLAN	73.14%	99,99%	14,48%
GENEVE	82.21%	99,99%	29,99%
NVGRE	99.94%	99,99%	99,97%

Figure 4 shows the worst, average and best prediction rate for each protocol. The colored rectangle marks the upper and lower quartile of the prediction.

Whereas the predicted detection rate of our trained model for VXLAN and GENEVE spread widely, the prediction rate of NVGRE remains nearly to the prediction rate of plain traffic. This can be explained by investigating the encapsulation structure in each protocol. VXLAN, as well as GENEVE, adds UDP and as a second addition, a protocol dependent header in front of the original network packet. Each UDP packet contains a checksum, which varies for every packet. This additional information results in a more different network protocol, which confuses the trained model, and leads to the reduced prediction rate.

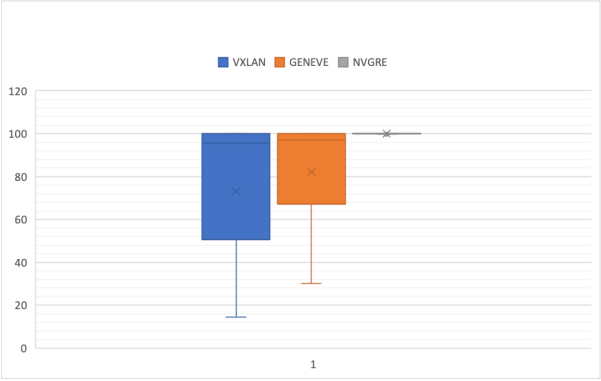


Fig. 4: Prediction distribution per protocol.

6 Conclusions

The use of deep learning algorithms provides an improved classification of network packets and detection of malicious or unwanted network traffic. But most of the research focus on plain network traffic without any encapsulation, and are therefore limited in their significance for modern and virtual environments. In this paper, we analyzed a model with a high detection rate of 99,98% when classifying non-encapsulated network traffic. Using this model for traffic classification of encapsulated network packets, the detection rate decreases to 70 to 80%. As a result, security implementations in modern networks using NVO have to consider the high dynamic and possible protocol changes inside the environment to reach or steady the needed detection rate.

The use of ICMP messages does not completely cover the real world, but it shows the risks of trusting deep learning models in highly dynamic networks. We use a static approach of network packet transformation, but virtual networks in operation include more possible shifts and different complex scenarios like user-customized networks. By this, customers are able to create their own networks in a provider environment, thus the provider does not have knowledge of these changes. As a result, the provider is unable to react timely. Our future research will focus on more real-world scenarios, supported by the analysis of well-known data-sets like CAIDA Anonymized Internet Traces 2016 Dataset[5] or KDDCUP99[6], either with or without encapsulated protocols.

Bibliography

[Ac10] Aceto, Giuseppe; Dainotti, Alberto; De Donato, Walter; Pescapé, Antonio: PortLoad: taking the best of two worlds in traffic classification. In: 2010 INFOCOM IEEE Conference on Computer Communications Workshops. IEEE, New York, NY, pp. 1–5, 2010.

[5] Online available at: `https://www.caida.org/data/passive/passive2016dataset.xml`
[6] Online available at: `http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html`

[Ac18] Aceto, Giuseppe; Ciuonzo, Domenico; Montieri, Antonio; Pescapé, Antonio: Mobile encrypted traffic classification using deep learning. In: 2018 Network traffic measurement and analysis conference (TMA). IEEE, New York, NY, pp. 1–8, 2018.

[Ac19] Aceto, Giuseppe; Ciuonzo, Domenico; Montieri, Antonio; Pescapè, Antonio: MIMETIC: Mobile encrypted traffic classification using multimodal deep learning. Computer Networks, 165:106944, 2019.

[AKO15] Al Khater, Noora; Overill, Richard E: Network traffic classification techniques and challenges. In: 2015 Tenth international conference on digital information management (ICDIM). IEEE, New York, NY, pp. 43–48, 2015.

[BAM10] Benson, Theophilus; Akella, Aditya; Maltz, David A: Network traffic characteristics of data centers in the wild. In: Proceedings of the 10th ACM SIGCOMM conference on Internet measurement. pp. 267–280, 2010.

[Fe03] Feinstein, Laura; Schnackenberg, Dan; Balupari, Ravindra; Kindred, Darrell: Statistical approaches to DDoS attack detection and response. In: Proceedings DARPA information survivability conference and exposition. volume 1, IEEE, New York, NY, pp. 303–314, 2003.

[GBC16] Goodfellow, Ian; Bengio, Yoshua; Courville, Aaron: Deep learning. MIT Press, Cambridge, MA, 2016.

[GGS20] Gross, Jesse; Ganga, Ilango; Sridhar, T.: , Geneve: Generic Network Virtualization Encapsulation. RFC 8926, November 2020.

[GW15] Garg, Pankaj; Wang, Yu-Shun: , NVGRE: Network Virtualization Using Generic Routing Encapsulation. RFC 7637, September 2015.

[He17] He, Hongmei; Watson, Tim; Maple, Carsten; Mehnen, Jörn; Tiwari, Ashutosh: A new semantic attribute deep learning with a linguistic attribute hierarchy for spam detection. In: 2017 International Joint Conference on Neural Networks (IJCNN). IEEE, New York, NY, pp. 3862–3869, 2017.

[KJ18] Kannojia, Suresh Prasad; Jaiswal, Gaurav: Effects of varying resolution on performance of CNN based image classification: An experimental study. Int. J. Comput. Sci. Eng, 6(9):451–456, 2018.

[Li19] Lim, Hyun-Kyo; Kim, Ju-Bong; Heo, Joo-Seong; Kim, Kwihoon; Hong, Yong-Geun; Han, Youn-Hee: Packet-based network traffic classification using deep learning. In: 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE, New York, NY, pp. 046–051, 2019.

[Lo20] Lotfollahi, Mohammad; Jafari Siavoshani, Mahdi; Shirali Hossein Zade, Ramin; Saberian, Mohammdsadegh: Deep packet: A novel approach for encrypted traffic classification using deep learning. Soft Computing, 24(3):1999–2012, 2020.

[Ma03] Mahoney, Matthew V: Network traffic anomaly detection based on packet bytes. In: Proceedings of the 2003 ACM symposium on Applied computing. ACM, New York, NY, pp. 346–350, 2003.

[Ma14] Mahalingam, Mallik; Dutt, Dinesh; Duda, Kenneth; Agarwal, Puneet; Kreeger, Larry; Sridhar, T.; Bursell, Mike; Wright, Chris: , Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks. RFC 7348, August 2014.

[MMS19] Mohammed, AysŞe Rumeysa; Mohammed, Shady A; Shirmohammadi, Shervin: Machine learning and deep learning based traffic classification and prediction in software defined networking. In: 2019 IEEE International Symposium on Measurements & Networking (M&N). IEEE, New York, NY, pp. 1–6, 2019.

[Po81] Postel, J.: Internet Control Message Protocol. STD 5, RFC Editor, September 1981. http://www.rfc-editor.org/rfc/rfc792.txt.

[SHA17] Shiddieqy, Hasbi Ash; Hariadi, Farkhad Ihsan; Adiono, Trio: Implementation of deep-learning based image classification on single board computer. In: 2017 International Symposium on Electronics and Smart Devices (ISESD). IEEE, pp. 133–137, 2017.

[SK20] Spiekermann, Daniel; Keller, Jörg: Impact of virtual networks on anomaly detection with machine learning. In: 2020 6th IEEE Conference on Network Softwarization (NetSoft). IEEE, New York, NY, pp. 430–436, 2020.

[SK21a] Spiekermann, Daniel; Keller, Jörg: Encapcap: Transforming Network Traces to Virtual Networks. In: 2021 IEEE 7th International Conference on Network Softwarization (NetSoft). IEEE, New York, NY, pp. 437–442, 2021.

[SK21b] Spiekermann, Daniel; Keller, Jörg: Unsupervised packet-based anomaly detection in virtual networks. Computer Networks, 192:108017, 2021.

[SK21c] Spiekermann, Daniel; Keller, Jörg: Unsupervised packet-based anomaly detection in virtual networks. Comput. Networks, 192:108017, 2021.

[SSE19] Sagduyu, Yalin E; Shi, Yi; Erpek, Tugba: IoT network security from the perspective of adversarial deep learning. In: 2019 16th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON). IEEE, New York, NY, pp. 1–9, 2019.

[SSJ21] Sadeghzadeh, Amir Mahdi; Shiravi, Saeed; Jalili, Rasool: Adversarial Network Traffic: Towards Evaluating the Robustness of Deep-Learning-Based Network Traffic Classification. IEEE Transactions on Network and Service Management, 18(2):1962–1976, 2021.

[Ta16] Tang, Tuan A; Mhamdi, Lotfi; McLernon, Des; Zaidi, Syed Ali Raza; Ghogho, Mounir: Deep learning approach for network intrusion detection in software defined networking. In: 2016 international conference on wireless networks and mobile communications (WINCOM). IEEE, 2016, pp. 258–263, 2016.

[Ts18] Tsai, Pang-Wei; Tsai, Chun-Wei; Hsu, Chia-Wei; Yang, Chu-Sing: Network monitoring in software-defined networking: A review. IEEE Systems Journal, 12(4):3958–3969, 2018.

[Vi19] Vinayakumar, R; Alazab, Mamoun; Soman, KP; Poornachandran, Prabaharan; Venkatraman, Sitalakshmi: Robust intelligent malware detection using deep learning. IEEE Access, 7:46717–46738, 2019.

[Wa18] Wang, Pan; Ye, Feng; Chen, Xuejiao; Qian, Yi: Datanet: Deep learning based encrypted network traffic classification in sdn home gateway. IEEE Access, 6:55380–55391, 2018.

[XLJ21] Xie, Guorui; Li, Qing; Jiang, Yong: Self-attentive deep learning method for online traffic classification and its interpretability. Computer Networks, 196:108267, 2021.

[Ze19] Zeng, Yi; Gu, Huaxi; Wei, Wenting; Guo, Yantao: $Deep-Full-Range$: a deep learning based network encrypted traffic classification and intrusion detection framework. IEEE Access, 7:45182–45190, 2019.

A pipeline for analysing image and video material in a forensic context with intelligent systems

Svenja Preuß[1] and Dirk Labudde[2]

Abstract: Shows like CSI seem to convey a certain view on the capabilities of forensic science based on the vast progress of digitalisation and the new technology, that goes along with it. But those depictions can be misleading and hardly represent a realistic reflection of reality. Nevertheless, this representation influences the public view on digital forensic analysis. This phenomenon is also known as the CSI effect. To present a more realistic view on practices in digital forensics, we want to introduce typical image and video analysis methods used in tackling real life forensic challenges and point to their capabilities as well as their limits. In this context an important area is image and video enhancement. With methods such as Super Resolution, images can be scaled up, their corresponding resolution gets enhanced and image noise can be reduced. During the subsequent image analysis, methods ranging from purely cognitive analysis to extraction of raw pixel values and various semantic information from the image or video, utilizing AI frameworks, are used. This allows for example to detect and analyse faces up to whole people, as well as objects in images or videos.

Keywords: forensic science; computer vision; AI frameworks

1 Introduction (CSI effect)

The profession of forensic scientist is gaining increasing interest among the general public [LS17]. Not least due to the numerous movies and shows, such as "CSI: Crime Scene Investigation" or its spin-offs "CSI: Miami", "CSI: NY" and other related shows. In these, scientific analyses and analysis techniques are increasingly moving in focus [LS17]. However, these analyses often hardly represent a realistic reflection of reality and are usually exaggerated, presented in a non-scientific style. Especially in the analysis of images and videos, sections can be enlarged repeatedly until even the smallest trace and detail is completely sharp visible. This results in a phenomenon which is heavily discussed in literature, the "CSI effect". [LS17] This phenomenon manifests itself in three different ways [PO05]. One, as alluded to earlier, is the increasing public interest and awareness of forensic science [PO05]. This is also reflected in the increasing demand for university programmes or vocational trainings in forensic science [LS17].

The attributed influence of the "CSI effect" on judges as well as other legislatives is

[1] University of Applied Sciences Mittweida, Department of Natural and Computer Sciences, Technikumplatz 17, 09648 Mittweida, Germany preuss2@hs-mittweida.de
[2] University of Applied Sciences Mittweida, Department of Natural and Computer Sciences, Technikumplatz 17, 09648 Mittweida, Germany labudde@hs-mittweida.de

doi:10.18420/inf2022_09

clearly more critical. The portrayal of forensic science in these shows creates an exaggerated expectation of courts for forensic analysis. Also, this expectation is reinforced by the displayed infallibility of the results of forensic analyses, creating a fear that all results will be accepted without question. [PO05]

Literature disagrees if the "CSI effect" has an actual impact on the legislative system or it is just a spurious correlation exaggerated by current findings. But it is undoubted that the shows have some influence on the point of view of forensic science in society, its methods and methodologies [LS17]. In order to draw a more realistic picture of digital forensics, especially with regards to digital multimedia forensics, some methods, which are applied in this field, will be presented here and evaluated on case studies.

2 Definition of forensic image processing and image analysis

In this section, the necessary terms for forensic image and video analysis shall be introduced and delimited from each other. Therefore, orienting on the work of Grimm F. [GR90], a distinction is made between **image analysis** and **image processing**. It should be noted that a video is nothing more than a sequence of images that are sequential in time, called frames [NI20]. Therefore, before any processing and analysis steps, the videos are separated into their individual frames. During image processing the original image is transformed so its information layout is more suitable for the respective analysis. A part of the image processing is the **image pre-processing**, whereby most of all the quality of the image should be improved. **Image segmentation**, likewise a subfield of the image processing, deals with the separation of the image into different image areas or its components. [GR90] In this way, relevant areas can be extracted and viewed independently of their surroundings. Image analysis, on the other hand, does not transform the image, but describes the content and semantic of the image in a certain way [GR90]. In the following, the methods of the individual field, which are used in daily work of a digital forensic scientist, will be displayed.

3 Image processing

3.1 Image pre-processing

One of the most frequent requests, which often appears in inquiries, is to improve the image or video quality. Often the material to be analysed is from security cameras located randomly near the crime scene, with the area of interest occupying only a small part of the image. Here we return to the example from the introduction, where a section of the image can be enlarged infinitively. An unlimited enhancement of image segments is not possible, since the resolution of an image also determines the information that can be displayed. Super resolution methods can be used in this context, to generate a higher-

resolution images from low-resolution ones, with the same content. However, it should be emphasized here that no new information can be gained, but only fill values can be estimated by linking existing information (interpolation). Two day to day used types of super resolution are on one hand basic interpolation and on the other hand methods that can be summarized as digital neural image enhancement technologies (DNET). The latter technologies not only increase the resolution, but also include other methods of image pre-processing, such as noise reduction or contrast enhancement.

Interpolation attempts to approximate unknown points of a function that lie between two or more known function values (samples) [FA14]. In terms of an image, the samples would each represent one pixel of the original image and the values to be determined would be the pixel values between these given pixels needed for magnification. Interpolation methods for calculating the intermediate values, which are mainly used here, are linear and cubic or bi-linear and bi-cubic. The difference between the former and the latter is that using the first two methods the interpolation is calculated in one dimension only and using the latter over two dimensions. Consequently, in bi-linear/bi-cubic interpolation, the nearest pixels are considered in both x and y directions. Usually, this two-dimensional interpolation is implemented by first interpolating in the x-direction and then in the y-direction [FA14]. As the name suggests, (bi-)linear interpolation is based on a linear function. In bi-linear interpolation the weighted average of the four nearest pixels is taken into account [FA14], all other pixels have no relevance. (Bi-)cubic interpolation is based on a third-degree polynomial function and considers the weighted average of the 16 nearest points [FA14], with closer points having a higher weighting [PMG13]. Cubic interpolation usually achieves better results than linear [FA14] [PMG13], but requires slightly more computation time [PMG13].

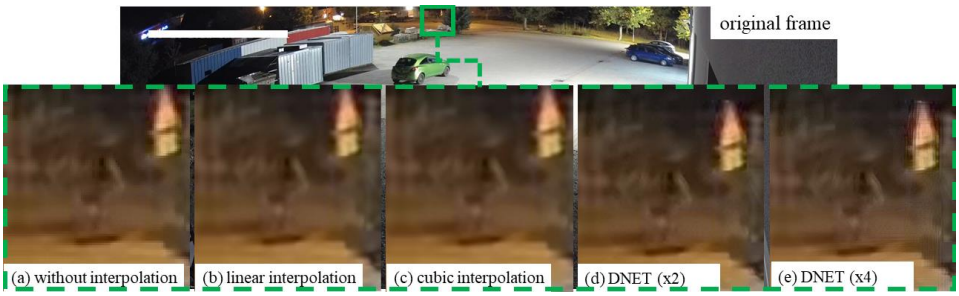


Fig. 1 Super resolution enlarged sections (a - e) from one frame (original frame) of the video footage from a security camera located near the scene of an attempted arson. The image sections show a person with a dog which had to be analysed.

Digital neural image enhancement technologies, on the other hand, are based on machine learning. They increase the resolution by learning the weights, based on data sets, to determine the information to be generated between the surrounding points. Consequently, this method generates the most probable weights of the surrounding pixel values based on the learned data. Currently, "neural-enhance", a free software for

enhancing images by Cardinale F. is applied [CA22]. In contrast to conventional Interpolation, these techniques have a probability of constructing information that was not originally present. As an example, a neural network trained on faces might generate a birthmark on the face even though there is actually no birthmark at that location, based on the trained data it is only very likely that there might be a birthmark.

Stacking is another way of enhancing visual perception, if several frames or images of the same scene or object are available. In the following, a distinction is made between **focus stacking** and **temporal stacking**. In focus stacking, individual images with differently sharp zones are combined to form a new image with a greater depth of field [KN19]. Temporal Stacking can also reduce noise [AT13] and merge overlapping image information by superimposing the individual frames. Temporal stacking is performed by several steps proposed by K. Atanassov [AT13]. Basically, a reference image has to be selected first, tending to choose the image with the highest information content and the lowest image noise [AT13]. Afterwards, key points are determined on the object to be stacked, which can be detected unambiguously and on the basis of which the frames can be aligned with each other [AT13]. Finally, the frames are combined with each other. Usually this is done by calculating the mean or average or by finding the median, for each pixel value across all frames. This can reduce random noise by producing constant signals more clearly.

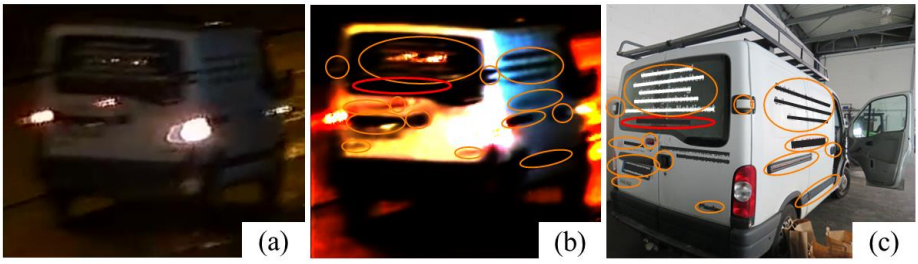


Fig. 2 (a): Reference image of a transporter captured by surveillance cameras at a local gas station. (b): Image stacked from six brightened frames of the transporter aligned using prominent keypoints of the transporter. (c): Image of the comparison transporter. The red and orange circles symbolize the individual feature for the correspondence analysis

However, stacking is only applicable if sufficient variable material is available. Furthermore, no non-existing information can be extracted here either. If material is only available in very poor resolution or quality, often insufficient information can be extracted. If temporal stacking does not provide adequate information, objects from different images whose similarity is to be investigated can also be compared by aligning them and fitting them to each other. Based on this, a later correspondence analysis (see section 3. step: "purely cognitive analyses") can be simplified. This derived form of stacking can be called **information merging**.

Stacking is already used in many different areas, for example biology, especially included microscopy [KN19]. Nevertheless, these methods can also be applied in a

forensic context.

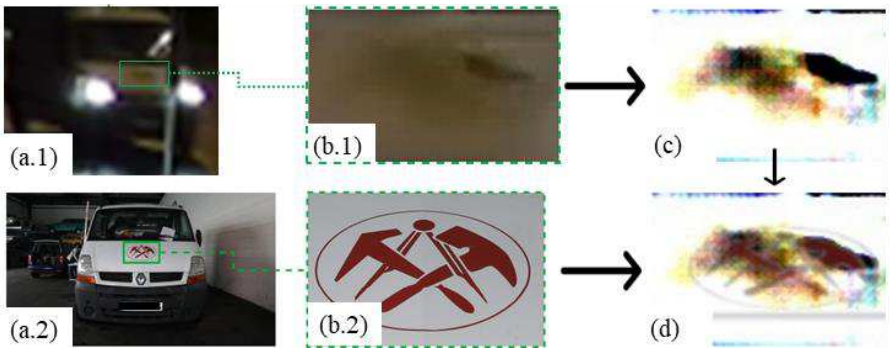


Fig. 3 Pipeline for information merging. (a): Transporters to be checked for matching. (b): Respective extracted area on the hood of the transporters (a.1) and (a.2). (c): Separately contrast and brightness adjusted extracted area on the bonnet of the transporter of (a.1). (d): Result of information merging.

Other improvements that can be used individually for the image and video material are the **adjustment of contrast** and **brightness** as well as **saturation**. Often a **white balance** is also necessary [BSS17]. **Colour normalization** is the adjustment of the colour histogram, which is a representation of various statistical information about the distribution of colour and brightness values [BSS17] in an image. By normalizing this histogram, colours can be balanced [IQ10] and a possible cognitively perceptible colour shift can be corrected. In addition to the above-mentioned stacking, **noise**, as described in [BSS17], can be **reduced** by smoothing the image, but always at the cost of reducing the level of detail.

3.2 Image segmentation

Often the area to be investigated in the video is only a very small part of the total captured area. In the case of recordings by surveillance cameras, the interest often lies additionally on the time stamp of the video, which is usually located in the upper area. To extract only the relevant information, non-essential areas of the image can be removed and the regions needed for analysis can be selected. If necessary, relevant information can be reassembled into an image afterwards. This allows better focus on the essential content and simplifies purely cognitive analyses that occur later.

4 Image analysis

4.1 "Purely cognitive analyses"

The method behind pure cognitive analysis is called **correspondence analysis**. In correspondence analysis, the aim is to find out, on a purely visual basis, whether an unknown object in an image or video matches a known object. For the correspondence-based method, in addition to the image of the unknown object, at least one separate image is needed that depicts either the same object or a very similar object. To compare the objects in both images, remarkable structures, which describe the examined object, have to be determined. These noticeable structures, also called features, representing characteristics for comparing the objects. If the same features are identified in matching locations when comparing the objects, a match can be assumed. When selecting the features, it should be checked previously whether the structures are not image artifacts, which can arise, for example, from an underlying compression process. This would falsify a correspondence analysis. To prevent this, if objects from video data are compared or more than one comparison image of the object is available, correspondence features can be tracked over several frames. A disadvantage of the cognitive analysis is that it is inherently subjective in nature due to the selection of features. In order to find corresponding features, an appropriate level of detail must also be available. Accordingly, this method is also limited by the quality of the image and video material.

4.2 "Analysis of raw pixel values"

Basically, a digital image is divided into a limited number of elements, called pixels, where each pixel has a unique, fixed location in the image and is assigned one or more finite, discrete values [CD11]. Mathematically, a digital image (I) is usually described as a matrix (1)

$$I = i(x, y), \tag{1}$$

wherein it is divided into image rows (L) with indexes x (x = 0, 1, 2 ... L-1) and image columns (R) with indexes y (y = 0, 1, 2, ..., R-1). The pixels already described above, also called picture element, each represent an element in I. i(x, y) consequently represents the value of a picture element at its corresponding location coordinates. [NI20]

These values can now be examined with respect to forensic problems and provide information about various properties of image objects. This is the case, for example, with **colour analysis**. For this purpose, the pixel values of an image area are averaged and the values are analysed. We have observed that a transformation of the colour values into the HSV colour space ([CD11]) is better suited for colour analysis.

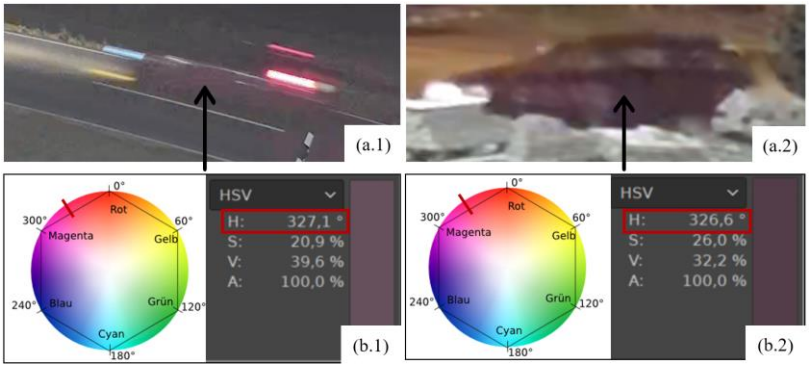


Fig. 4 Colour analysis of two vehicles captured by surveillance cameras of a car dealership near the crime scene (a). Both vehicles were observed to have a colour between magenta and red, which is illustrated by the colour circles and the red lines (b).

However, the lighting situation can have a great influence on the captured colour values and thus lead to irreversible shifts of the values in the colour space. In addition, if an object is further away from the camera, the number of available pixels is reduced, which means that all colours within the available pixels, for the object, are represented mixed. A similar phenomenon can be observed when working with highly compressed images, if several pixels are encoded and stored together during compression. In that case, an accurate colour analysis can be difficult or even impossible if the object to be analysed is represented by only a few pixels.

4.3 Extraction of semantic information

Machine learning, among other techniques, is used to extract semantic information from the images. For example, free AI frameworks can be used to extract and match faces.

To be able to analyse complete humans, another framework has been established, the neural network based "OpenPose". This is used to predict key points on the human body based on 2D image information. "OpenPose" is the first real-time multi-person system that can simultaneously recognize human body, hand, face and foot "keypoints" on individual images. The process is divided into several steps. After receiving a 2D image as input, the network generates a 2D confidence map for each body part position. In addition, 2D vector fields for part affinity fields (PAFs) are computed. These represent the directed connection from each keypoint to the adjacent keypoints of the same person. Now, based on the confidence maps, all graphs connecting different, possibly adjacency keypoints are created. Then, weaker links are removed using the PFAs. As a final step, the 2D positions of the anatomical keypoints for each person in the image are predicted from the information obtained in the previous steps. [CA18]

This framework is particularly useful in combination with externally generated

information. Thus, this can be used to match the motion or skeletal apparatus of a known person with an unknown person in the video. To do this, a digital skeleton (Rig) first has to be derived, also based on "OpenPose", from the underlying image information, which is obtained from individual images of a photogrammetric scan. Rigging as a working technique in 3D animation fundamentally refers to the construction of a skeleton or rig based on bones to determine how individual parts of a 3D model can be moved and can be applied analogously to a 2D context. By merging and comparing this "OpenPose"-Rig of a known person with the information obtained through "OpenPose" of the keypoints of an unknown person in an image or video, a conclusion can be made about whether the unknown person is the known person or not. In order to enable such a comparison, not only the known person himself, but also the surrounding areas as well as the crime scene must be captured photogrammetrically and transferred into a 3D reference model. In the resulting 3D space, measurements and comparisons of the rigs can be performed.



Fig. 5 Comparison of the digital skeleton of a known person, generated by "OpenPose", during an extended measurement service (a) with a digital skeleton of an unknown person, captured by a surveillance camera of a gas station (b).

This procedure builds on the assumption that the skeleton and thus also the digital skeleton of a person is unique [BE22].

The problem with all these methods is the usually very poor quality of the video and image material. This can lead to deviations in the predictions. In addition, the detection of faces seems to have difficulties in face recognition when other strong patterns dominate in the face area, which additionally obscures parts of the face, as well as when the level of detail of the face is too low.

In conclusion, many methods shown in shows like CSI are based on real methods, but their competence is highly exaggerated. The methods described here represent a more realistic state of forensic image and video analysis.

5 Case studies

The process of applying the individual steps does not necessarily follow a strict pattern. The individual steps can be applied repetitively and sometimes in a different order.

All methods of Super Resolution explained can be compared on the basis of the first case to be discussed. This involved an attempted arson. Video footage from a car dealership surveillance camera, which happened to be near the crime scene, showed a person with a dog (Fig. 1). The task was to analyse both the person and the dog. As can be seen in Fig. 1 original frame, the person occupies only a small part at the edge of the field of view of the surveillance camera. In order to be able to perform a cognitive analysis, the first task was to extract the relevant area by image segmentation and to enlarge it according to Super Resolution methods (Fig. 1 (a)-(d)). In this way, for example, a size classification of the dog could be performed and the basic clothing of the person could be described. In addition to the analysis of the person with the dog, two vehicles should also be analysed, which were captured by two cameras of the car dealership while passing by (Fig. 4 (a)). Here, the assumption arose that they could be the same or similar vehicles. In the process of these analyses, the vehicles were examined for their colouration with the help of colour analysis (Fig. 4 (b)), once colour normalization has taken place. Both vehicles were observed to have a colour between magenta and red (Fig. 4 (b) red lines).

The second case to be considered is a preliminary investigation on suspicion of extortionate kidnapping from several apartments. The aim was to determine whether the passing transporter (Fig. 2 (a), (b)) recorded by a surveillance camera at a gas station, with apparent advertising spaces, was the comparison transporter provided by the police (Fig. 2 (c)). Again, the key image sections were first extracted using image segmentation. Then, six frames could be matched and aligned based on keypoints of the van. Based on the cumulated image, the extraction of registrable features for the correspondence analysis (Fig. 2 orange and red circles) could take place. Particularly noticeable were individual lettering and applications (such as handles) on the back and side of the transporter (Fig. 2 (b)). Once sufficient features have been extracted, a comparison can be made with the reference transporter (Fig. 2 (c)). Here, not only matching features are noted, but also deviating conspicuous features. Another very characteristic feature is a logo on the bonnet of the comparison transporter, which can be recognized from the high-resolution images provided (Fig. 3 (a.1)). At the same time, several frames could be extracted from a video of the surveillance camera, on which frame-spanning dark pixels can also be registered on the bonnet of the transporter (Fig. 3 (a.1)). After an initial extraction of the relevant areas (Fig. 3 (b)) and a separate adjustment of contrast and brightness of the frame section (Fig. 3 (c)), both extracted regions could be aligned and adjusted to each other through an information matching (Fig. 3 (d)).

For matching people, as in the case of a robbery series, "OpenPose" can be used. In this robbery series, a few suspects were available (Fig. 5 (a)) as well as a surveillance video of a gas station, which had recorded the robbery (Fig. 5 (b)). Extended measurements of the subjects were taken, with the goal of creating Rigs for all individuals with the help of "OpenPose" (Fig. 5 (a) color-coded lines). "OpenPose" was also able to generate the respective Rigs for the people that were captured by the surveillance camera based on the 2D information (Fig. 5 (b) colour-coded lines). After creating a 3D model of the crime scene based on a photogrammetric scan, the Rigs could be matched. In this way,

an assignment of the known suspects to the persons recorded during the crime could take place.

Bibliography

[At13] Atanassov, K. et al.: Temporal image stacking for noise reduction and dynamic range improvement. In: Kennedy S. et al. (Hg.): Multimedia Content and Mobile Devices. IS&T/SPIE Electronic Imaging, Burlingame, California, USA, 2013

[BE22] Becker, S. et al.: COMBI: Artificial intelligence for computer-based forensic analysis of persons. In: KI - Künstliche Intelligenz, (in press) 2022

[BSS17] Bühler, P.; Schlaich, P.; Sinner, D.: Digitale Fotografie. Fotografische Gestaltung - Optik - Kameratechnik. Springer, Berlin, Heidelberg, Germany, 2017

[CA18] Cao, Z. et al.: OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2018.

[CA22] Cardinale, F.; et al.: ISR. https://github.com/idealo/image-super-resolution. accessed: 31/05/2022

[CD11] Chanda, B.; Dutta Majumder, D.: Digital image processing and analysis. Second edition. PHI Learning Private Limited. Delhi, 2011.

[FA14] Fadnavis, S.: Image Interpolation Techniques in Digital Image Processing: An Overview. In: Int. Journal of Engineering Research and Applications (10), pp. 70–73, 2014

[GR90] Grimm, F.: Expertensysteme für den Einsatz von Subroutinenpaketen. Am Beispiel eines Expertensystems für Bildverarbeitung. Springer, Wiesbaden, Germany, 1990

[IQ90] Iqbal, K.: Enhancing the low quality images using Unsupervised Colour Correction Method. In: IEEE International Conference on Systems, Man and Cybernetics (SMC), Istanbul, Turkey. pp. 1703–1709, 2010

[KN19] Knop, D.: Schärfentiefe nach Maß. In: Biol. Unserer Zeit 49 (1), pp. 48–57, 2019

[LS17] Labudde, D.; Spranger, M.: Forensik in der digitalen Welt. Moderne Methoden der forensischen Fallarbeit in der digitalen und digitalisierten realen Welt. Springer, Berlin, Heidelberg, Germany, 2017

[NI20] Nischwitz, A. et al.: Bildverarbeitung. 4. Auflage. Springer, Wiesbaden, Heidelberg, Germany, 2020

[PMG13] Patel, V.; Mistree, K.; Gopalbhai, C.: A Review on Different Image Interpolation Techniques for Image Enhancement. In: International Journal of Emerging Technology and Advanced Engineering 3 (12), pp. 129–133, 2013

[PO05] Podlas, K.: The CSI Effect": Exposing the Media Myth. In: Fordham Intellectual Property, Media & Entertainment Law Journal 16, pp. 429, 2005

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 119

Collecting Identifying Data for Re-Identification of Mobile Devices carried at a Crime Scene using Wi-Fi Routers

Daniel Vogel[1], Markus Krämer[2]

Abstract: Mobile devices such as smartphones constantly probe the surrounding wireless medium for providing services to users, thus leaving traces that, when analysed, can link their presence to a crime. Those passive observations, combined with actively tricking devices to cooperate with foreign networks, can yield an attractive device profile. Though past research explores independent methods for collecting single identifiers of devices, a holistic approach for data seizure and utilization for law enforcement is missing. As devices carried by criminals are not willing to share identifiers with a victim's home network, we combine collection processes for device-identifying data with exploiting viable wireless communication protocols. Using the presented approach allows to build a system supporting law enforcement with evidence that certain mobile devices were present at a crime scene and thus may give clues on further investigations on suspects.

Keywords: Wi-Fi; MAC-Address; Cellular; IMSI; Evidence; Law Enforcement

1 Introduction

More than 100,000 thefts by burglary of a dwelling are registered only in Germany every year. Still, the crime clearance rate is rather low, fluctuating between 16-18%. [Po]
Therefore, provisions to reach a higher clearance rate are desirable which may in addition lead to less burglaries happening. We aim at providing strategies for using home routers as alarm equipment and furthermore, as a collector of re-identifying information of burglars' technical equipment. While there are existing technologies for localising Wi-Fi devices via home routers [BP00, AHP17], these are out of the scope of this paper and are assumed to be working and given. Goal of this paper is creating a collection of Device Identifying Data (DID) types collectable via home routers, a classification of their usefulness in crime clearance and as these techniques collide with anti-tracking technologies, collecting workarounds against those techniques to allow a re-identification of devices brought by burglars. The rest of the paper is organized as follows:
Sect. 2 includes a motivation and our use case. Sect. 3 comprises an introduction of DID from the Wi-Fi and the cellular domain while Sect. 4 introduces collection processes for the DID. Related Work is discussed in Sect. 5. The paper will be concluded in Sect. 6.

[1] Universität Bonn, Informatik IV, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany vogel@cs.uni-bonn.de
[2] Universität Bonn, Informatik IV, Friedrich-Hirzebruch-Allee 8, 53115 Bonn, Germany kraemerm@cs.uni-bonn.de

doi:10.18420/inf2022_10

2 Background and Use Case

When breaking into homes, police intelligence shows that burglars often bring communication devices in order to keep in contact with their accomplices. When considering non-organized groups, it can be assumed that amateur burglars might bring their phone just in a lack of awareness.
This presents an detection opportunity as communication devices such as phones transmit signals even when currently not used that can be detected by any nearby receiver that is set up accordingly. The goal is to utilize the commonly existing home router as that receiver as Wi-Fi routers are ubiquitously available in most homes and provide a commonly used networking interface for mobile phones already. This scenario is visualised in figure Fig. 1.

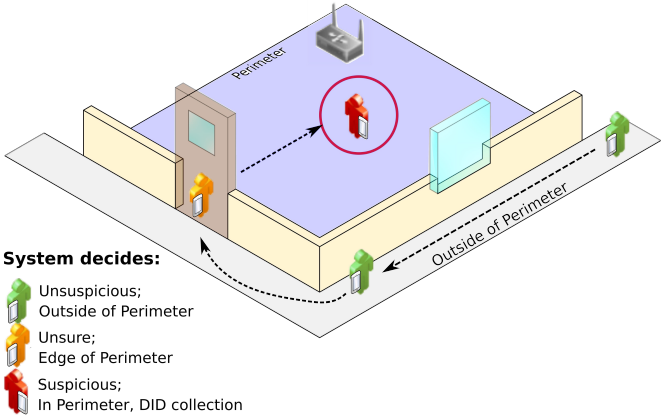


Fig. 1: Use case: Brought devices of different suspicion levels as detectable by the home router.

As soon as the router can eavesdrop packets from a mobile device it evaluates whether the device is located inside a certain perimeter or not. Therefore, three different states exist. Devices localised outside the perimeter are considered as unsuspicious. At the edge of the perimeter it might be unsure whether the device has entered the perimeter or not. Thus, the device will be observed more intensively. In case the the device has surely entered the perimeter, the system alarms the user and collects as much DID as possible.
However, as we assume that mobile phones of burglars are not set up to willingly connect to the victim's home network in order to collect their data, some detection and collection strategy needs to be employed. We present a strategy that looks to collect DID for a brought non-cooperative phone by its sent packets solely through the usage of a home router.
In this paper the following terminology will be used:

- **Target Device (TD)** denotes a mobile device, most commonly a mobile phone, that is brought by a malicious intruder and will be the target for re-identification.
- **Defensive Router (DR)** denotes the Commercial-Off-The-Shelf (COTS) router a home has already installed and is used to collect information on the TD.

Note that devices need to be located inside a defensive perimeter to be considerable as TDs. The federally founded german research project WACHMANN is currently researching exactly that scenario and thus we consider the requirement as satisfactory for the scope of this paper [WA].

2.1 Viable Wireless Technologies

Mobile phones use a number of wireless communication technologies depending on different factors such as brand, vendor, series and more. Since most wireless technologies try to serve different requirements, they vary in both their technical specification, protocols, and transmitted data. When deciding on whether a technology is suitable for our use case, the capabilities of the chosen receiver as well as the capabilities of the technology itself have to be considered. As the home router shall be utilized as the DR, we restrict the technically available wireless technologies. Not all wireless technologies commonly used by mobile phones are supported by COTS routers, such as Bluetooth or cellular. In short, the wireless communication technology Wi-Fi provides both the most promising approaches as well as the highest level of compatibility between existing home routers and the most commonly expected TD, being mobile phones.
Mobile phones, when not connected to a Wi-Fi network, will only use certain frames [IE21]. These frames comprise management frames used for network discovery or general control frames. Since these frames are of limited use for device re-identification as presented in Sect. 3.1, it may prove useful to entice the TD to send more and preferably different frames. Applying enticing strategies accordingly will allow the DR to receive different Wi-Fi frames in certain situations, thus allowing for further information gathering. In Sect. 4 we present different strategies, a DR can employ to receive packets from a TD.

2.2 Criteria for Device Identifying Data

Concerning Wi-Fi packets hovering around, such as probe requests, many data types can be extracted from these. Even more data types can be collected through forcing more and tailored communication with foreign Wi-Fi devices. However, not all of these data types can be considered as DID. While specific DID are given in Sect. 3 the following catalogue of criteria should give an imagination of features needed for data types being DID:

- **Technical traceability:** There must be a technical possibility to trace collected DID. This includes reproducability at other locations at a later point of time.
- **Legal traceability:** Due to the strict data privacy guidelines of the german law BDSG [Re14] and the european GDPR [Ge] the use of DID may be restricted to specific DID or specific use cases.

- **Uniqueness:** DID must be unique or at least collision free in the considered area.
- **Re-identifiability:** The given DID allows a re-identification of wireless devices and is helpful in detecting criminals.
- **Unforgeability:** DID should be hard to forge.

3 Device Identifying Data

Some DID are useful by design and used for exactly that purpose, to identify and address unique devices in a heterogeneous wireless network environment. Other information are not designed specifically to allow identification but support other services. These, however, can be utilized as DID when viewed in combination with each other and in context of the use case.
DID are used to solve two distinct tasks. Firstly, in the context of an ongoing burglary, DID are used to attribute packets to unique devices. This is important as the DR receives packets from potentially many devices in a certain time frame, many of whom are not carried by burglars. Correctly attributing the sent packets from the TD to their original sending device allows for a complete overview on what has been sent by this device during the whole duration the DR was able to observe it. Not only is this information crucial on gathering intelligence on the device and its behaviour, being able to tell individual devices apart by their packets assists packet-based device positioning. Secondly, once sufficient DID were collected for a TD, they can be used to re-identify that TD at a later time. Correctly re-identifying a mobile device and linking it to a crime scene, where its attributing DID were collected, is the main goal to achieve as presented in this paper.

3.1 Device Identifying Data from Wi-Fi

Since we selected Wi-Fi as the wireless communication technology to focus on, Wi-Fi-based DID will be shortly summarized now. We distinguish between *direct* DID and *indirect* DID. It must be noted that there are more information extractable from Wi-Fi packets than listed here, such as Channel State Information (CSI), or Time of Flight (ToF). We omitted these as they either have uses primarily for device positioning, which is not within the scope of this paper, or do not majorly assist as device identifiers in a way that a COTS router can make use of.

Direct DID are DID that allow device identification directly, meaning that they can be extracted directly from the received packets without the need of further interpretation or context. For our use case we consider Media Access Control (MAC) addresses and Information Elements (IEs). In Wi-Fi, MAC addresses are used for addressing individual devices in a wireless networking scenario and are meant to be a unique device identifier by design [IE21]. Thus, in their originally intended use, they serve as a direct DID for

our purpose as they allow to directly attribute packets containing certain MAC addresses to individual devices. In practice, modern mobile phones tend to utilize MAC address randomization to enhance the privacy of these devices [Va16, Fe21]. As MAC addresses may provide complications and can therefore not solely be relied on for device re-identification, other approaches for DID are considered.
Wi-Fi probe requests and responses contain what are called IEs. IEs contain information used for different uses, ranging from supported rates, requested Service Set Identifiers (SSIDs), up to Wi-Fi features specific information such as Wi-Fi Protected Setup (WPS) data. Research shows that IEs allow to generate fingerprints for packets, which can allow to attribute packets correctly to individual devices, especially when combined with an analysis of sequence numbers. [Va16, Ma17]
Due to fabrication inaccuracies mobile devices could be identified on specific characteristics during Wi-Fi transmissions, such as minor deviations in the used transmission frequencies [US07]. We consider DRs however to not be sufficiently sensitive receivers to utilize physical device profiles as DID.

Indirect DID are not by themselves of great use, often because they are not unique per device, but can provide additional insights which may amplify the usefulness of direct DID in re-identification. We consider information on time, vendor, Operating Systems (OSs), SSIDs, User Equipment (UE) usage, and logical device profiles as indirect DID.
Time information can proof useful in the context of both sketching the process of an ongoing burglary as well as assisting the law enforcement in giving a burglary context between other burglaries. Mobile phones or UEs regularly probe the surrounding wireless medium for available networks, which are identified by their SSIDs. Whether a UE probes using SSIDs or using wildcard probe requests affects both their power efficiency but also their privacy [Da18, KA05]. Using knowledge about SSIDs in combination with services such as WiGLE [bau] allows for further tracking opportunities.
UEs are further shown to send packets differently depending on their usage. For example different mobile phones wait for varying time slots between sending packets based on whether the display is locked or unlocked, or whether their Wi-Fi is enabled or not [Wa13, MCT17, Fr15]. Using this knowledge, it is possible to infer the usage of a UE based on the sent packets and their amount within a time frame.

3.2 Device Identifying Data from the Cellular Domain

From the cellular domain, three different DID types have been identified which strictly speaking are not DID of the mobile phone but identifier of the SIM card of the mobile user. However, this information still is really helpful to get information about the mobile user or the International Mobile Equipment Identity (IMEI) of the used device, as pairs of International Mobile Subscription Identity (IMSI) and IMEI are stored by the cellular provider for up to seven days. This allows law enforcement to study the history of the DID IMEI through the IMSI. Cellular networks also use identifiers such as the TMSI and GUTI,

where it is unclear, whether these identifiers are stored in a way that they can be useful to our use case. [ET21]

4 Methods for Data Collection

A holistic approach for data seizure and utilization for law enforcement given our use case (c.f. Sect. 2) is yet missing. To formulate such an approach, we compiled various collection methods for DID presented in past research and enriched this compilation with known attacks on commonly uses Wi-Fi communication protocols. The three main strategies for DID collection are:

1. Passively receiving the packets the TD is sending ordinarily.
2. Actively engaging the TD to increase the amount and variety of packets sent by it.
3. Actively engaging the TD, tricking it to try to connect to a network, the DR is claiming to provide access to.

These strategies can be used in combination or be layered into different modes of operation, the DR can employ given its burglary detection certainty, especially when taking privacy into account.

4.1 Passive Data Collection

Passively receiving packets sent by the TD and all other devices nearby means that no further active communication shall be issued by the DR other than what it usually would do to provide its service to nearby devices. Usually this means that the TD would only engage minimally with the DR by occasionally sending Wi-Fi frames, such as probe requests. Even this minimally invasive strategy will yield DID as present in probe requests albeit in a limited supply. Observables are MAC addresses, although likely randomized for probe requests, IEs usable for fingerprinting, physical and logical device profiles, and most indirect DID, depending on how the TD is used and set up. It is very much possible to observe further transmissions by the TD, if it is actually communication e.g. with some nearby network to further collect more information.

4.2 Active Data Collection

As alluded to earlier, receiving more and different packets from the TD assists in creating a more reliable fingerprint. Past research shows, how certain packets like data [AA20] or Ready To Send (RTS) [Wa13] packets can be utilized to force responses like Acknowledgement

(Ack) or Clear To Send (CTS) packets even from non-cooperative devices. Abedi et al. suspect that the required response times as defined by the Wi-Fi standard is not sufficient to check the validity of said packets [AA20].
Certain Wi-Fi features such as *WPS* or *Passpoint* (also called *Hotspot 2.0* in *Wi-Fi Vantage* [Wi20]) may allow for additional communication between two unconnected devices, as these features are specifically designed to aid in establishing connections. A DR could be set up to transmit Passpoint-enabled and WPS-enabled packets aiming to collect information from TD supporting these Wi-Fi features.

4.3 Router-based Attacks for Data Collection

The third approach to gathering more information on the TD focuses on exploiting vulnerabilities of Wi-Fi protocols. While MAC-randomization usually is enabled for sent packets in an unconnected state, a successful connection requires the mobile device to be addressable by the network. As such, mobile devices will either revert to using their original Wi-Fi MAC address or a MAC address that is rolled for this network specifically, when attempting to associate to a network. In the latter case, we call that MAC address a session MAC address and mobile phones do maintain an internal database of network SSIDs and corresponding session MAC addresses.
The primary goal is to trick the TD to try to connect to a network that the DR is advertising, using so called association attacks. These association attacks have been discussed thoroughly as a way to attack MAC randomization, as the internal memory of the TD would directly allow linking the TD to any network it has a database entry for. It is important to note that for most known attacks these will not result in a fully established Wi-Fi connection with a malicious Access Point (AP). As the desired information can however be extracted just from the attempt of a connection establishment, a fully established connection is not actually required.
There are several attacks presented in past research that aim to either hijack an existing Wi-Fi connection or trick a mobile device to connect to an AP, revealing its sensitive information. Once such an association attacks succeeds and the TD tries to establish a connection to the DR, additional DID can be collected to improve both the fingerprinting precision and thus the likelihood of a successful re-identification of the TD. Furthermore, in [OBH17] O'Hanlon et al show, how Extensible Authentication Protocol (EAP) can be exploited to build a Wi-Fi Based IMSI Collector (WBIC). Using this technique, we were able to successfully extract the IMSI of TD in certain scenarios using an SSID that is pre-configured on certain SIM cards used in Germany.

4.4 Resulting Unified Collection Approach

The presented approaches can be unified into a unified collection approach. The DR is set up to monitor Wi-Fi packets passively. Wi-Fi packets that it receives will be used for

device fingerprinting using direct and indirect DID. These fingerprints will be used to differentiate between packets from different devices and can be used for any implemented indoor positioning approach to identify TDs. Once, a TD has been identified, meaning a device has been spotted inside the protected perimeter, both active data collection and data collection attacks can be utilized to gather even more information to improve the fingerprint for the TD. This fingerprint may then be used to re-identify the TD.
Assuming a fingerprint has been generated that is sufficient in re-identifying the TD correctly, the device must first be available for comparison. Though this is not focus of this paper, some directions include direct intervention by law enforcement, cell site analysis (if an IMSI has been collected), or various MAC or SSID tracking techniques. Generated fingerprints could further be used to link crimes with each other, if identical fingerprints have been generated on different occasions.

5 Related Work

From our research it has become clear that Wi-Fi as a technology is presenting viable leads to police investigation that is not well integrated into common investigation mechanisms, primarily because there is no reliable infrastructure to capture these DID. Since the research in the field of digital forensics is vast and we cited many relevant publications, we selected some examples to present, just to highlight further work in nearby use cases.
The analysis of wireless traces for device detection has been discussed e.g. in [Bu20], where they discuss that criminals may set up a system to detect nearby mobile devices carried by law enforcement officers in order to warn themselves against an upcoming police raid. Here too, previously collected DID are utilized to correctly detect mobile devices nearby, though the intention is malicious.
Wi-Fi device identification is also relevant to network administrators to prevent malicious participants to wreak havoc in their network, potentially exploiting an open network, as Yu et al present in [Yu20]. Here, passively collected packets are fed into a deep learning framework to assess identity anomalies and allow device classification.

6 Discussion and conclusions

In this paper we presented a holistic approach in analyzing DID of intruders' mobile phones leading to new ways in crime clearance through digital forensics. We examined Wi-Fi as a viable technology and its receivable packet types. In addition, we featured a catalogue of criteria for classifying data types as DID. Using this catalogue we worked out the DID from the Wi-Fi and cellular domain. Furthermore we presented collection processes for these and workaround for existing problems, such as MAC randomization.
Concerning the practical use an effective countermeasure might be to shut the smartphone down or even do not bring it to burglaries. While this would neutralize our DR, on the other hand the technical possibilities for communication in burglary groups would decrease.

As an effective approach device-free localization [Zh21, Da19] could detect intruders without a need for carrying Wi-Fi devices, but is not able to collect DID then. There is a race between anti-tracking techniques such as MAC randomization which affects crime clearance and countermeasures such that further research is needed in the future to keep a working system. We stress, that the presented approaches must be evaluated clearly with regards to precision, presented criteria, as well as conformity with applicable (privacy) laws in the context of burglaries.

Acknowledgement

The work on this paper has been funded by the German Federal Ministry of Education and Research (BMBF) for the project WACHMANN (WLAN-based recording of characteristics of mobile devices near a crime scene for alarming and tracing property crimes) under grant number 13N15359.

Bibliography

[AA20] Abedi, Ali; Abari, Omid: WiFi Says"Hi!"Back to Strangers! In: Proceedings of the 19th ACM Workshop on Hot Topics in Networks. pp. 132–138, 2020.

[AHP17] Ali, Muhammad Usman; Hur, Soojung; Park, Yongwan: Locali: Calibration-free systematic localization approach for indoor positioning. Sensors, 17(6):1213, 2017.

[bau] bobzilla; arkasha; uhtu: , WiGLE: Wireless Network Mapping. `https://wigle.net`. Accessed: 2021-05-20.

[BP00] Bahl, Paramvir; Padmanabhan, Venkata N: RADAR: An in-building RF-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Conference on computer communications. Nineteenth annual joint conference of the IEEE computer and communications societies (Cat. No. 00CH37064). volume 2. Ieee, pp. 775–784, 2000.

[Bu20] Bug, Steffen: Wenn der Schutzmann nicht mal klingelt - Über das automatische Erkennen von und Warnen vor der Staatsgewalt vor des Straftäters Haustür. 06 2020.

[Da18] Dagelić, Ante; Perković, Toni; Vujatović, Bojan; Čagalj, Mario: SSID oracle attack on undisclosed Wi-Fi preferred network lists. Wireless Communications and Mobile Computing, 2018, 2018.

[Da19] Dang, Xiaochao; Si, Xiong; Hao, Zhanjun; Huang, Yaning: A novel passive indoor localization method by fusion CSI amplitude and phase information. Sensors, 19(4):875, 2019.

[ET21] ETSI: . TS 123 003 V16.5.0 Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); LTE; 5G; Numbering, addressing and identification (3GPP TS 23.003 version 16.5.0 Release 16), January 2021.

[Fe21] Fenske, Ellis; Brown, Dane; Martin, Jeremy; Mayberry, Travis; Ryan, Peter; Rye, Erik C: Three Years Later: A Study of MAC Address Randomization In Mobile Devices And When It Succeeds. Proc. Priv. Enhancing Technol., 2021(3):164–181, 2021.

[Fr15] Freudiger, Julien: How talkative is your mobile device? An experimental study of Wi-Fi probe requests. In: Proceedings of the 8th ACM Conference on Security & Privacy in Wireless and Mobile Networks. pp. 1–6, 2015.

[Ge] General Data Protection Regulation. https://gdpr-info.eu/. Accessed: 2022-06-01.

[IE21] IEEE: IEEE Standard for Information Technology–Telecommunications and Information Exchange between Systems - Local and Metropolitan Area Networks–Specific Requirements - Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std 802.11-2020 (Revision of IEEE Std 802.11-2016), pp. 1–4379, 2021.

[KA05] KARMA Attacks Radioed Machines Automatically. https://web.archive.org/web/20210401071739/http://theta44.org/karma/index.html, 2005. Accessed: 2021-04-01.

[Ma17] Martin, Jeremy; Mayberry, Travis; Donahue, Collin; Foppe, Lucas; Brown, Lamont; Riggins, Chadwick; Rye, Erik C.; Brown, Dane: , A Study of MAC Address Randomization in Mobile Devices and When it Fails, 2017.

[MCT17] Matte, Célestin; Cunche, Mathieu; Toubiana, Vincent: Does disabling Wi-Fi prevent my smartphone from sending Wi-Fi frames? PhD thesis, Inria-Research Centre Grenoble–Rhône-Alpes; INSA Lyon, 2017.

[OBH17] O'Hanlon, P.; Borgaonkar, R.; Hirschi, L.: Mobile Subscriber WiFi Privacy. In: 2017 IEEE Security and Privacy Workshops (SPW). pp. 169–178, 2017.

[Po] Polizeiliche Kriminalstatistik. Edited: 2021-02-20.

[Re14] Recht, G.: Bundesdatenschutzgesetz (BDSG). G. Recht, 2014. ISBN:978-1-500-10002-5.

[US07] Ureten, Oktay; Serinken, Nur: Wireless security through RF fingerprinting. Canadian Journal of Electrical and Computer Engineering, 32(1):27–33, 2007.

[Va16] Vanhoef, Mathy; Matte, Célestin; Cunche, Mathieu; Cardoso, Leonardo; Piessens, Frank: Why MAC Address Randomization is not Enough: An Analysis of Wi-Fi Network Discovery Mechanisms. In: ACM AsiaCCS. Proc of ACM AsiaCCS, Xi'an, China, May 2016.

[WA] WACHMANN: WLAN-basierte Aufzeichnung von Charakteristiken tatortnaher mobiler Endgeräte zur Alarmierung und Nachverfolgung von Eigentumskriminalität.

[Wa13] Wang, Wei; Joshi, Raj; Kulkarni, Aditya; Leong, Wai Kay; Leong, Ben: Feasibility study of mobile phone WiFi detection in aerial search and rescue operations. In: Proceedings of the 4th Asia-Pacific workshop on systems. pp. 1–6, 2013.

[Wi20] Wi-Fi Alliance: . Wi-Fi CERTIFIED Passpoint Technology Overview, December 2020.

[Yu20] Yu, Lingjing; Luo, Bo; Ma, Jun; Zhou, Zhaoyu; Liu, Qingyun: You Are What You Broadcast: Identification of Mobile and IoT Devices from (Public) WiFi. In: 29th USENIX Security Symposium (USENIX Security 20). USENIX Association, pp. 55–72, August 2020.

[Zh21] Zhang, Yunwei; Wang, Weigang; Xu, Chendong; Qin, Jie; Yu, Shujuan; Zhang, Yun: SICD: Novel single-access-point indoor localization based on CSI-MIMO with dimensionality reduction. Sensors, 21(4):1325, 2021.

Mehr Frauen in die IT – Wie bekommen wir das hin?

Yes she can! How to encourage girls to choose a career path in computer science

Janki Dodiya, Cornelia Heinisch, Claudia Hess, Sibylle Kunz, Sandra Rebholz, Inga Schlömer, Adrienne Steffen, Silke Vaas[1]

Abstract: Women are still underrepresented in studies and in job positions in the fields of STEM. This paper proposes a study and a lecture concept as measures to increase the number of women in Computer Science careers. The study aims to better understand barriers, drivers and influencing factors that affect young women's decision for or against a career in computer science. A lecture series on computer science is proposed, specifically designed for girls and young women to reduce fears and prejudices and increase the visibility of women in IT. This will also provide further research data that will enable adjusting to the requirements of the target group based on the research outcomes. With this article, the authors would like to initiate discussion on how universities can better gear their offerings to a female target group.

Keywords: Women in STEM, computer science careers, female-oriented lecture concept

1 Introduction and State of the Art

Science, technology, engineering, and mathematics (STEM) are deemed as highly significant to any national economy and important fields of action of a government's high-tech strategy. An equal gender-based participation within these fields is an essential condition for the sustainable and future-oriented development of our society, because “diversity” is the key to creativity, productivity, and innovation [Ki05] [BU09].

Women are still underrepresented in education and in academic positions in the field of computer science. In winter term 2020/21, within the field of computer science, there were only 25.735 female students (19%) enrolled in studies of computer science in German Universities as opposed to 108.030 male students (81%) [St21]. According to current research, the reasons preventing girls and women to major in computer science are manifold: Cultural stereotypes, gender biases, and microinequities (subtle discriminatory behaviors) evoke a missing sense of belonging to these fields [CH15]; lack of female role models and mentors that provide support and encouragement; low level of self-efficacy, engagement, and perceived low practical relevance; missing gender-sensitivity in computer science classrooms and in the learning design [SOL20]. To overcome these barriers, there are many recommendations and initiatives on how to promote girls and women in computer

[1] IU Internationale Hochschule, Juri-Gagarin-Ring 152, 99084 Erfurt
Fachbereich IT & Technik: janki.dodiya@iu.org, cornelia.heinisch@iu.org, claudia.hess@iu.org, sibylle.kunz@iu.org, sandra.rebholz@iu.org, inga.schloemer@iu.org; silke.vaas@iu.org
Fachbereich Wirtschaft & Management: adrienne.steffen@iu.org

doi:10.18420/inf2022_11

science. These include measures like integrating technology and computer science education in the early stages of the curriculum (e.g. [GI19]), offering online coding platforms for all ages (e.g. Scratch, code.org), improving career counseling and orientation options (school and non-school activities) [Dr19] as well as setting up girls-only initiatives to provide creative and inspiring learning environments (e.g. Girls' Day, Hackathons for Girls). Current research also investigates how to make science and higher education more inclusive and accessible. Organizations such as „500 Women Scientists" (https://500womenscientists.org/) specifically address women and gender minorities in STEM and launch initiatives such as supporting *mothers in science*, *women scientists with disabilities*, and conducting so-called Wikipedia Edit-a-thons in order to increase the visibility of women in STEM on Wikipedia.

2 Study on career choice drivers and barriers

Key factors for increasing STEM career interest are also linked to increasing the interest in these fields among elementary, middle and high school students [VW12], [Bu21]. While both girls and boys take math and science courses in school, and despite the similarities in their achievements, fewer women than men pursue STEM majors. This transition from high school to college/university seems to be a critical moment when many young women turn away from a STEM career path [HI10]. Based on these findings and existing experiences, further research is needed on how to encourage and support young women in deciding on a career path in computer science. Current work to perform a study seeks to answer the following research questions:

- What are barriers, drivers and key situations that affect girls' and young women's decision to enter the field of computer science?
- Who are the influencers for a career path in computer science?
- Which offers are already being used for career orientation and decision support?

The target group is female secondary school students (15 - 19 years) who are close to finishing their secondary school years and will be deciding their career path soon. Data collection will take place in May and June 2022 via an online access panel from a market research institute. An online panel is particularly suitable to find a representative sample for special samples e.g., a particular age group because the socio-demographic data is available upfront [DB16]. The targeted sample size will be approximately 1500 students. Screening criteria are used in the beginning of the questionnaire so that about 60% of the sample will have an interest in STEM topics. This is done to evaluate whether the answers of students with an interest in STEM topics and those without differ significantly. In addition, 60% of the sample should also be interested in higher education study to be able to evaluate whether students who consider an apprenticeship in a STEM area differ from those who seek a degree in higher education. A structured survey is used which combines mainly quantitative questions with a few qualitative questions to gain insights into the motivations, barriers, influencers and information sources for career orientation.

3 Lecture Series

To actively offer a personal and interactive way of informing about computer science, a lecture series for the target group was designed. It addresses current topics from computer science and related fields of study in the scope of six virtual lectures of 60min each. In a combination of theory and case studies, it will be explained how computer science impacts the daily lives of each of us. The objectives are to explain what computer science is, remove fears and prejudices, and increase the visibility of women in IT. The lectures are given exclusively by female professors from IU International University of Applied Sciences, who also talk about their own careers and are available to answer participants' questions. In addition, they feature female scientists and entrepreneurs who are active in the respective fields. By using interactive elements like online quizzes, observations or short surveys, additional data about motivation, self-awareness and image of STEM studies will be collected. The topics cover a broad range of fields of computer science and each has a significant message to be conveyed:

Computer science is *universal* in its application: We look at a typical day in a life of the girls and check, where there is "IT inside". What kinds of human computer interfaces do we use and how? What kind of data do we provide and what happens to it?

Computer science is no magic*:* What does a computer scientist do? What is programming? What is an algorithm? What does it feel like to work in IT as a female professional? What female strengths are important for this?

Computer science is *creative:* A computer scientist thinks out of the box and solves real-life problems in innovative ways. How can software help you to explore your digital creativity?

Computer science is *pure communication:* What is communication? Why do we need rules and protocols? How do computer networks work? How do we communicate with computers? And how does communication help us to develop better software?

Computer science *changes the world:* Computer science helps to optimize processes and products. What aspects of life would you like to improve?

Computer science has many interesting *research fields*. Which fits best *for me*? What's the difference between Applied Computer Science, Business Informatics, Artificial Intelligence, Cyber Security, Data Science, Digital Humanities, Robotics and Medical computer science? Where would you like to work and why?

As part of the planned series of lectures, the requirements of the target group will be continuously surveyed. From the study results and the additional findings collected during the lecture series, measures can be derived to help increase the representation of women in the study of Computer Science. Furthermore, the measures for addressing and providing orientation to the target group will be evaluated. In this way, the effects of the individual measures can be reviewed and adjusted if necessary.

4 Conclusion

The development of high-quality higher education for girls and women expands their career opportunities. This paper considered the underrepresentation of women in STEM and presented a proposal of a study and a lecture concept as measures to increase the number of women in computer science careers. The study aims to better understand the critical barriers, drivers, and key situations as well as influencing factors and relevant orientation offers that affect girls' and young women's decision for or against computer science as a major. The results of this study will give insights into target group needs and for designing appropriate course offerings in computer science. Also, the findings will be directly incorporated into the design of the lecture concept presented, aiming at positive associations and the reduction of possible barriers regarding a study in computer science.

Bibliography

[Bu21] Bundesministerium für Bildung und Forschung - BMBF (2021): Bekanntmachung. Available online: https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2021/08/2021-08-19-Bekanntmachung-Mint.html

[BU09] Bell, S. & O'Halloran, K. & Saw, J. & Zhao, Y. (2009). Women in Science: Maximising productivity, diversity and innovation.

[CH15] Corbett, C.; Hill, C.: Solving the Equation. The variables for Women's Sucess in Engineering and Computing. AAUW, US, Washington D.C., 2015.

[DB16] Döring, N., & Bortz, J. (2016). Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften (5. edition). Springer, p. 297.

[Dr19] Dreisiebner, G.: Berufsfindungsprozesse von Jugendlichen. Eine qualitativ-rekonstruktive Studie. Springer Fachmedien Wiesbaden GmbH, 2019.

[GI19] Gesellschaft für Informatik (GI): Kompetenzen für informatische Bildung im Primarbereich. Empfehlungen der Gesellschaft für Informatik e.V., Bonn, 2019.

[HI10] Hill, C. & Corbett, C. & Rose, A. (2010). Why So Few? Women in Science, Technology, Engineering, and Mathematics. American Association of University Women.

[Ki05] King, J. (2005): Benefits of women in science: American Association for the Advancement of Science (308)

[St21] Statistisches Bundesamt. Anzahl der Studierenden im Fach Informatik in Deutschland nach Geschlecht in den Wintersemestern von 1998/1999 bis 2020/2021. https://de-statista-com.pxz.iubh.de:8443/statistik/daten/studie/732331/umfrage/studierende-im-fach-informatik-in-deutschland-nach-geschlecht/, Stand: 09.04.2022

[VW12] Valla, J.M., Williams, W. (2012): Increasing achievement and higher-education representation of under-represented groups in science, technology, engineering, and mathematics fields: A review of current K-12 intervention programs. In: Journal of women and minorities in science and engineering 18 (1).

Alex wird Informatikerin - Eine Geschichte von Studierenden für Jugendliche zur Berufsinformation

Nadine Dittert[1], Ira Diethelm[2]

Abstract: Das Berufsbild der Informatikerin ist für Jugendliche wenig bekannt oder attraktiv. In dem Seminar „Gendergerecheter Informatikunterricht" haben Informatik-Lehramtsstudierende eine Geschichte für Jugendliche verfasst, in der vier Informatikberufsbilder vorgestellt werden. Sie stellen dabei eine 17-jährige Schülerin in den Mittelpunkt, die sich mit verschiedenen Personen trifft, um sich ein Bild von verschiedenen Möglichkeiten als Informatikerin zu machen. Die Geschichte soll als Anregung für Informatiklehrkräfte dienen, jungen Menschen verschiedene Informatikberufe vorzustellen. Ebenso können Hochschulen diese Geschichte an Schulinformationstagen nutzen, um dort über Studieninhalte oder Abschlussmöglichkeiten zu informieren. Die Geschichte steht unter einer cc-Lizenz zur weiteren Verbreitung und auch zur Bearbeitung zur Verfügung, was explizit erwünscht ist.

Keywords: Informatikerin; Beruf; Geschichte; Schülerin; Schüler; Gender

1 Einleitung und Begründung

Trotz vieler Bemühungen und steigender Zahlen sind Studentinnen in der Informatik nach wie vor selten. An der Universität Oldenburg lag der Anteil weiblicher Studierender im Fachbachelor bzw. -master im Wintersemester 2021/22 bei 12% bzw. 18% [UO22], im bundesweiten Schnitt mittlerweile knapp über 20%, wobei im Lehramt Informatik ein etwas höherer Anteil weiblicher Studierender von etwa 30% zu verzeichnen ist (Daten von 2020, [Ko22]).

Die Gründe für das unausgeglichene Verhältnis von Männern und Frauen in der Informatik sind vielfältig. So werden stereotype Vorstellungen von Männern und Frauen, fehlende Vorbilder, das negative Image der Informatik und auch die geringere Selbstwirksamkeitserwartung von Mädchen gegenüber Jungen oft als Begründungen genannt [BSW08], [Be14], [BSA18], [Va19]. Dies wurde, so ein Erklärungsversuch, vor allem in westlichorientierten Ländern in den 1980er Jahren durch die Vermarktung des C64 an Jungen und die bis heute andauernde stereotype Werbung unterstützt [He14]. Stockfotos und ihre entsprechende Verbreitung durch voruteilsbehaftete Journalist:innen und Schulbuchautor:innen zeichnen ein ausschließlich männliches Bild. Frauen werden außerdem, falls sie denn auf illustrierenden

[1] Universität Potsdam, Didaktik der Informatik, An der Bahn 2, 14476 Potsdam, Deutschland, dittert@uni-potsdam.de
[2] Carl von Ossietky Universität Oldenburg, Didaktik der Informatik, Uhlhornsweg 84, 26129 Oldenburg, Deutschland, ira.diethelm@uol.de

doi:10.18420/inf2022_12

Bildern im Kontext Informatik zu sehen sind, eher in assistierenden Rollen oder als Opfer von Cyberkriminalität gezeigt. Weiterhin gehört die Verwendung nicht-gendergerechter Sprache zu den Gründen, warum bereits Mädchen einige Berufe für sich nicht in Betracht ziehen [VH15].

2 Die Entstehung der Idee

Im Seminar „Gendergerechter Informatikunterricht", das an der Universität Oldenburg erstmalig im Sommersemester 2021 für Lehramtsstudierende im Master of Education stattfand, haben sich die Lehramtsstudierenden nicht nur mit den verschiedenen Ursachen der geringen weiblichen Beteiligung in der Informatik beschäftigt, sondern auch konkrete Lösungsvorschläge diskutiert und erarbeitet. Nach der theoretischen Auseinandersetzung mit den Themen Stereotype, Vorbilder, gendergerechte Sprache und Selbstwirksamkeitserwartung haben vier Studierende beschlossen, eine Geschichte zu verfassen, die das Ziel hat, das Berufsbild der Informatikerin für Jugendliche möglichst realitätsnah und vielfältig zu beschreiben. Aufgrund der vorhergegangenen Besprechungen flossen die im Seminar thematisierten Aspekte in die Erzählung ein. Die Idee war es dabei vor allem, entgegen Klischees in heutigen Kinderbüchern zu arbeiten. Häufig werden dort Berufe implizit, beispielsweise die Eltern der Protagonistin, oder explizit, wie in Büchern zur Berufsvorstellung, noch immer stereotyp, etwa als Pilot oder Krankenschwester, präsentiert. Da die Studierenden selbst in ihrer späteren Berufslaufbahn mit Schülerinnen und Schülern der Sekundarstufen I und II arbeiten werden, wurde dieses Alter als Zielgruppe für die Geschichte definiert. Als Inhalt aus der Lebenswelt der Jugendlichen eignete sich das Thema Berufswahl hervorragend. Die Idee zur Geschichte „Alex wird Informatikerin" war geboren.

3 Die Geschichte „Alex wird Informatikerin"

Die Geschichte „Alex wird Informatikerin" erzählt auf etwa acht A4-Seiten von der 17-jähigen Alex, die die 12. Klasse besucht und sich mit der Wahl ihres späteren Berufs auseinandersetzt. Dazu trifft sie sich zunächst auf ein Gespräch mit ihrer Informatiklehrerin Frau Hopper, die ihr über ihren eigenen Beruf und ihren Weg dorthin, insbesondere vom Studium, berichtet. Nach dem Gespräch weist Frau Hopper Alex auf die Berufsorientierungsmesse in der kommenden Woche hin. Dort lernt Alex die Anwendungsentwicklerin Karlie und den Medizininformatiker Konrad kennen. Beide erzählen von ihrem Berufsalltag, der durch unterschiedliche Projekte und Arbeiten im Team geprägt ist. Karlie entwickelt in einem Team ein Spiel und Konrad berichtet, wie er Computerprogramme entwickelt und trainiert, um Tomographien menschlicher Schädel zu analysieren. Zuletzt nimmt Alex' Vater sie mit zu seiner Arbeit und stellt ihr dort eine Fachinformatikerin für Systemintegration vor, die ihr ebenfalls eindrücklich von ihrem Berufsalltag erzählt und ihn ihr zeigt. Am Ende der Geschichte steht für Alex fest, dass sie irgendetwas mit Informatik machen möchte.

Die Geschichte wurde unter Begleitung der Dozentin, jedoch von den Studierenden selbst als gemeinsame Seminaraufgabe, verfasst. Die Illustration der Geschichte übernahm eine externe Zeichnerin. Hauptziel ist es zunächst, Jugendlichen verschiedene Berufsbilder in der Informatik zu zeigen. Dabei haben die Studierenden ihre eigenen Erfahrungen aus dem Lehramtsstudium eingebracht und sich darüber hinaus mit Berufsbildern befasst und diese dargestellt. Die weiteren Aspekte, die im Seminar theoretisch behandelt wurden, sind im Bezug auf die Geschichte diskutiert und reflektiert worden und flossen in verschiedenen Ausprägungen ein. Bezüglich der Darstellung der Personen sollten keine stereotype Vorstellungen bedient werden. So wurde das Verhältnis von weiblichen zu männlichen Informatiker:innen bewusst umgedreht. Von vier Berufsbildern werden drei von Frauen dargestellt, lediglich einer der Berufe wurde durch einen Mann vertreten, der in seiner Darstellung ebenso wenig wie die anderen Personen irgendein Klischee eines Informatikers oder einer Informatikerin bedient. Der Aspekt der Vorbilder wurde vorwiegend durch das explizite Zeigen weiblicher Berufstätiger umgesetzt, die in dem Feld aktiv arbeiten. Gleichzeitig wurde bei der Namensgebung der Personen, die Alex trifft, sowie der Schule auf die Namen berühmter Informatiker:innen zurückgegriffen, die ebenso als Vorbild dienen können. Die Bezüge zu den realen Informatiker:innen werden selbst nicht explizit thematisiert. Jedoch werden Anreize geschaffen, in dem beispielsweise die Lehrerin sagt „[...] wenn du das nächste Mal vor dem Schultor stehst und etwas Zeit hast, achte mal auf den Namen der Schule! Das Ada-Lovelace-Gymnasium heißt nicht grundlos so! Vielleicht findest du ja heraus, warum gerade diese Frau als Namensgeberin genommen wurde!" Die Namensgebung der Personen wird in den abschließenden Hinweisen für die Lehrkraft am Ende der Geschichte aufgelöst. Sie soll Lehrkräften als Anregung dienen, Erfindungen von Informatikerinnen im Unterricht zum Gesprächsanlass zu nehmen. Im Bezug auf die verwendete Sprache haben sich die Studierenden nach langen Überlegungen auf die Darstellung des Doppelpunkts mit einer gesprochenen Lücke, dem Glottisschlag, geeinigt. Diese Entscheidung fiel nicht leicht, da sie von vielen nicht als optimal empfunden wird, wurde aber schließlich mehrheitlich befürwortet, da Alternativen wie beispielsweise das Ausschreiben der männlichen und weiblichen Form nicht alle Geschlechter abbildet, was den Studierenden jedoch wichtig war. Obwohl die Protagonistin weiblich ist, wurde ein geschlechtsneutraler Name verwendet, um auch hier mögliche Assoziationen mit Klischees zu vermeiden. Der Aspekt der Selbstwirksamkeit war schwierig in der Geschichte explizit umzusetzen, da sie kein direktes Erfolgserlebnis des Zielpublikums ermöglicht. Lediglich Ermutigung für den Beruf wird hier implizit vermittelt, mit dem Ziel, die Selbstwirksamkeitserwartung insbesondere junger Frauen zu erhöhen.

4 Fazit und Ausblick

Die Entwicklung der Geschichte „Alex wird Informatikerin" im Seminar „Gendergerechter Informatikunterricht" diente zunächst der Sensibilisierung zukünftiger Informatiklehrkräfte für die Thematik „Gender und Informatik". Sie bot den Studierenden Anlass, sich aktiv mit Aspekten zu befassen, die zuvor theoretisch erarbeitet wurden, indem sie direkt in einem

Schreibprozess umgesetzt wurden. Als fertiges Produkt soll die Geschichte jedoch auch bei der Zielgruppe ankommen. Sie steht als Open Content mit cc-Lizenz Interessierten zur Verfügung[3] und kann und soll im Schulunterricht gelesen, auf Hochschulinformationstagen genutzt und bei weiteren Aktivitäten für Schulen verteilt werden. Dazu wird zusätzlich ein Poster erstellt, das in Schulen oder auf Veranstaltungen aufgehängt werden kann. Die Wirkung der Geschichte auf Schülerinnen und Schüler wurde bisher nicht evaluiert. Die Umsetzung einer entsprechenden Studie befindet sich in Planung.

Literaturverzeichnis

[Be14] Beyer, Sylvia: Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. Computer Science Education, 24(2-3):153–192, Juli 2014.

[BSA18] Berg, Tessa; Sharpe, Alexander; Aitkin, Emma: Females in computing: Understanding stereotypes through collaborative picturing. Computers & Education, 126:105–114, 2018. Publisher: Elsevier.

[BSW08] Bath, Corinna; Schelhowe, Heidi; Wiesner, Heike: Informatik: Geschlechteraspekte einer technischen Disziplin. In (Becker, Ruth; Kortendiek, Beate, Hrsg.): Handbuch Frauen- und Geschlechterforschung, S. 821–833. VS Verlag für Sozialwissenschaften, Wiesbaden, 2008.

[He14] When Women Stopped Coding. https://www.npr.org/sections/money/2014/10/21/357629765/when-women-stopped-coding, [Online; letzter Zugriff am 14.05.2022].

[Ko22] Datentool Studium. https://www.kompetenzz.de/service/datentool-studium, [Online; letzter Zugriff am 14.05.2022].

[UO22] Fachsemesterübersicht: Studierende (Fallstatistik) nach Studiengang und Fachsemester im WiSe 2021/22. https://uol.de/fileadmin/user_upload/referatplanung/Akademisches_Controlling/01_Studium_Lehre/02_Studierende/Stud-SUE2_FSUE_20212_211207_EV.pdf, [Online; letzter Zugriff am 14.05.2022].

[Va19] Vainionpää, Fanny; Kinnula, Marianne; Iivari, Netta; Molin-Juustila, Tonja: Girls' Choice - Why Won't They Pick IT? In: Proceedings of the 27th European Conference on Information Systems (ECIS). Stockholm & Uppsala, Sweden, 2019.

[VH15] Vervecken, Dries; Hannover, Bettina: Yes I Can!: Effects of Gender Fair Job Descriptions on Children's Perceptions of Job Status, Job Difficulty, and Vocational Self-Efficacy. Social Psychology, 46(2):76–92, März 2015.

[3] https://cloud.uol.de/s/ai3BeMoZN4WJ58y

STEM Girls – Eine Workshopreihe zur Förderung von Frauen und nicht-binären Personen in STEM-Berufen

Carolin Neumann[1], Clarissa Sabrina Arlinghaus[2]

Abstract: Die STEM GIrls Eventreihe der Jungen Gesellschaft für Informatik dient der Nachwuchsförderung von zukünftigen STEM Frauen* und wurde 2022 erstmals durchgeführt. Die Reihe besteht aus neun digitalen Workshops für weibliche und non-binäre Personen, in denen die Bereiche Mindset, berufliche Orientierung und Bewerbungen behandelt werden. Mit 367 Anmeldungen erlangte STEM GIrls ungewöhnlich viel öffentliche Aufmerksamkeit verglichen mit anderen Events der Jungen Gesellschaft für Informatik. Dadurch konnte der Bekanntheitsgrad der Jugendorganisation gesteigert werden und neue Mitglieder akquiriert sowie bestehende, inaktive Mitglieder reaktiviert werden. Durch Netzwerkarbeit, Vorstellung weiblicher Vorbilder, Empowerment und Unterstützung bei der Jobsuche soll der Anteil von Frauen und nicht-binären Personen in STEM Berufen langfristig erhöht werden. Besonderer Fokus wird dabei auf berufliche Perspektiven in der Informatik gelegt. Aufgrund der positiven Resonanz ist eine Fortführung der Eventreihe für die nächsten Jahre angedacht.

Keywords: WomenInSTEM, Nachwuchsförderung, SoftSkills, Workshops

1 Einleitung

Das englische Akronym STEM steht für Science, Technology, Engineering und Mathematics. STEM stellt somit einen Sammelbegriff für eine Vielzahl an Berufen dar, die für unsere gegenwärtige und zukünftige Gesellschaft von hoher Bedeutung sind. Trotz der hohen Relevanz entscheiden sich aufgrund geschlechtsspezifischer Hürden nur wenig junge Mädchen und Frauen für einen Beruf in diesen Bereichen [DS14; RM18]. Geschlechtsbezogene Stereotype, Erwartungen und Normen sowie eine zu geringe Übereinstimmung mit persönlichen Zielen führen dazu, dass sich Mädchen bereits in der Kindheit und Jugend von STEM-Fächern abwenden. Durch den hohen Männeranteil und fehlende weibliche Vorbilder fühlen sich viele junge Frauen in STEM-Bereichen als Außenseiterin, weshalb sie seltener STEM-Studiengänge ergreifen und sie häufiger vorzeitig beenden. Im Berufsleben werden viele Frauen mit geschlechtsspezifischen Vorurteilen konfrontiert, die ihre Einstellung und Beförderung, die Bewertung ihrer Arbeit sowie die Vereinbarkeit von Beruf und Familie betreffen [DS14]. Wissenschaftlich gibt es verschiedene Theorien darüber, ob eine geschlechtsorientierte Sozialisierung, Personen im gleichen Alter oder Vorurteile für den geringen Frauenanteil in STEM-Disziplinen verantwortlich sind [RM18].

[1] Junge Gesellschaft für Informatik, carolin.neumann@gi.de
[2] Junge Gesellschaft für Informatik, clarissa.arlinghaus@gi.de

doi:10.18420/inf2022_13

Als Praxisimplikation lässt sich festhalten, dass Netzwerkarbeit, Mentoring-Programme und weibliche Vorbilder wirksame Mittel darstellen, um Hemmnisse abzubauen und Frauen in STEM-Berufen zu unterstützen [DS14]. STEM-Bewegungen, pädagogische Strategien, Erfahrungen sowie das Fördern von Positivität und Neugier sind ebenfalls zur Verringerung der Geschlechterungleichheit geeignet [RM18]. Daran anknüpfend spricht die STEM GIrls Workshop-Reihe der Jungen GI gezielt weibliche und non-binäre Personen an, stellt weibliche Vorbilder vor, ermöglicht Vernetzung und möchte so langfristig einen Beitrag für mehr Geschlechterdiversität in den Bereichen Naturwissenschaften, Technik, Ingenieurwissenschaften, Informatik und Mathematik leisten. Zu diesem Zweck werden neun digitale Workshops zu Themen wie Softskills, selbstbewusstem Auftreten und Bewerbungen durchgeführt. Die genauen Titel und Beschreibungen der einzelnen Workshops können dem Anhang B entnommen werden.

2 Planung und Durchführung

Für die Organisation waren Mitglieder der Jungen GI zuständig. Die erste Idee der Eventreihe entstand Oktober 2021. Die Planung der einzelnen Themen begann unmittelbar danach und setzte den Fokus auf Mindset, berufliche Orientierung und Bewerbungstraining, woraus sich die drei Themenschwerpunkte der Eventreihe ergaben. Ein Netzwerkevent wurde für den Start der Workshopreihe vorgesehen und ermöglichte den Teilnehmenden, sich mit anderen STEM Girls* auszutauschen. Dies sollte eine gute Grundlage für die kommenden Veranstaltungen schaffen sowie Ängste und Unsicherheiten abbauen. Nach dem Kick-off Event sollten Workshops folgen, die die Teilnehmenden durch die Förderung verschiedener Soft Skills empowern sollten. Die Teilnehmenden sollten dazu ermutigt werden, sich neuen Herausforderungen zu stellen. Nach den ersten vier Veranstaltungen zur Mindset-Bildung wurde ein Abschnitt zur beruflichen Orientierung geplant. Der Orientierungsanteil soll den Teilnehmenden einen Einblick in verschiedene STEM-Berufsbereiche geben. Als letzter Abschnitt sind zwei Workshops zum Thema Bewerbungen vorgesehen, die den Teilnehmenden im Rahmen von Bewerbungstrainings ihren eigenen Wert am Bewerbungsmarkt verdeutlichen. Dies soll verhindern, dass Frauen und nicht-binäre Personen ihren Wert unterschätzen und infolgedessen aus Unsicherheit STEM-Jobs mit ungerechter Bezahlung oder schlechtem Arbeitsklima annehmen.

November 2021 wurden die Planung der Themen, Speaker:innen und Termine vorgenommen und die ersten Anmeldeseiten auf der Website der Jungen GI angelegt. Die Speaker:innen wurden über vorhandene Netzwerkstrukturen (u. a. BYTE Challenge und MINTORME) oder via LinkedIN gefunden. Für jedes Event sollten ein bis drei Gäste eingeladen werden. Dabei handelte es sich um Frauen im Alter von 30 bis 50 Jahren, die den Teilnehmenden als Vorbild und Inspiration dienen sollten. Ab Dezember 2021 wurden die Marketingmaterialien vorbereitet und die Feedbackbögen erstellt. Ebenso wurde hier das Social Media Marketing gestartet. Vereinbart wurden zwischen drei und zehn Social Media Posts auf mindestens zwei bekannten Plattformen wie

Instagram, Twitter oder LinkedIN. Darüber hinaus sollte in Rundmails und Newslettern der Jungen GI auf die Eventreihe aufmerksam gemacht werden. Die meisten Registrierungen sind auf den Mailverteiler der Jungen GI zurückzuführen, der überwiegend studentische Mitglieder der Jungen GI zu einer Teilnahme motivieren konnte. Durch Social Media erhöhte sich die Außenwirksamkeit, wodurch auch andere Interessierte erreicht werden konnten. Die Beiträge auf LinkedIN erwiesen sich als besonders wirksam, da sie von anderen Organisationen geteilt wurden und so die Reichweite der Eventreihe maßgeblich vergrößerten.

Alle Events wurden über Zoom durchgeführt. Ansonsten stand es den Referent:innen frei, die Veranstaltungen individuell zu gestalten, was zu verschiedenen Workshop-Formaten führte. In den meisten Fällen wurden Breakout Session und interaktive Aufgaben gewählt, aber auch von Folien unterstützte Vorträge mit anschließender Fragerunde waren Teil der Eventreihe. Die Moderation der Events wurde von Carolin Neumann, Annabella Kadavanich und Valeria Tisch übernommen, drei engagierten Mitgliedern der Jungen GI.

3 Teilnehmende

Im Durchschnitt registrierten sich die Teilnehmer:innen für 2.16 Veranstaltungen. Insgesamt konnten so 367 Anmeldungen von 170 Personen für die bereits stattgefundenen Workshops verzeichnet werden. Je nach Event schwankte die Anzahl der Anmeldungen zwischen 16 und 62 (siehe Tabelle 1).

Eventkategorie	Eventname	Registrierte Teilnehmende
Mindset	Netzwerken	38
Mindset	Digital You	31
Mindset	Selbstbewusst auftreten	48
Mindset	Komfortzone verlassen	42
Mindset	Stärken finden	16
Orientierung	Freelancing	39
Orientierung	IT Jobs	34
Bewerbung	Gehaltsverhandlungen meistern	57
Bewerbung	Bewerbungen rocken	62
	Gesamt	367

Tab. 1: STEM GIrl Events und dessen Registrierungen

Die zum Teil deutlich geringere Anzahl an Anmeldungen bei einigen Veranstaltungen ist auf ein zu spätes Veröffentlichen der entsprechenden Events zurückzuführen. Ein Großteil der Teilnehmenden registrierte sich bereits nach der ersten Rundmail über diese

Eventreihe zu allen derzeit veröffentlichten Events, die sie interessierten. Später veröffentlichte Events konnten nicht die gleiche Anzahl an Registrierungen erreichen.

Das Durchschnittsalter betrug über alle Workshops hinweg 25.34 Jahre. Unter den Teilnehmenden waren Studierende (60.7%), Berufstätige (17.9%) und Schüler:innen (16.1%) vertreten. 5.4% gaben als Berufsstatus „Sonstiges“ an. Die Mehrheit der Teilnehmenden war GI-Mitglied (62.5%) oder denkt über eine GI-Mitgliedschaft nach (12.5%). Insgesamt nahmen deutlich mehr weibliche (98.2%) als nicht binäre Personen (1.8%) teil.

33.3% der Teilnehmer:innen wurden durch den Mailverteiler auf die Workshopreihe aufmerksam. Die anderen Befragten gaben an, dass sie von Kontakten (15.2%), dem Newsletter (15.2%), Social Media (9.1%), der Website (3.0%) oder auf anderem Wege (24.2%) von der Workshopreihe erfahren haben.

4 Evaluation und Ausblick

Nach Abschluss der Workshopreihe konnte auf freiwilliger Basis ein Fragebogen ausgefüllt werden. Die Resonanz auf die Eventreihe fiel positiv aus (siehe Anhang A), sodass eine Fortsetzung unter Leitung von Clarissa Arlinghaus für die nächsten Jahre geplant ist. Gelobt wurden u. a. die offene und lockere Atmosphäre sowie die Möglichkeit, sich zu vernetzen. 90.0% gaben an, dass sie mit der Workshopreihe insgesamt zufrieden waren. Die Weiterempfehlungsrate fiel mit einem Net Promoter Score von 8.50 hoch aus. Als besonders hilfreich wurde die Workshopreihe für Berufseinsteiger:innen eingeschätzt.

Abgefragt wurde auch, welche von zehn Themenvorschläge in kommenden Workshops bearbeitet werden sollen. Für die Neuauflage der Workshopreihe wurden sich insbesondere die Themen Selbstbewusstsein (66.7%), IT Jobs (66.7%), Gehalt (66.7%), Netzwerken (61.9%) und Ziele (61.9%) gewünscht (Mehrfachnennung war möglich). Wir werden dies bei der Planung für STEM GIrls 2.0 berücksichtigen.

Abschließend bedanken wir uns bei allen beteiligten Personen und bei Christine Regitz, der Präsidentin der Gesellschaft für Informatik e.V. sowie Schirmherrin der STEM GIrls.

Literatur

[DS14] Dasgupta, N.; Stout, J. G.: Girls and Women in Science, Technology, Engineering, and Mathematics: STEMing the Tide and Broadening Participation in STEM Careers. Policy Insights from the Behavioral and Brain Sciences 1/14, S. 21–29, 2014.

[RM18] Reinking, A.; Martin, B.: The Gender Gap in STEM Fields: Theories, Movements, and Ideas to Engage Girls in STEM. Journal of New Approaches in Educational Research 7/14, S. 148–153, 2018.

A Feedback

Auf die Frage „**Warum nimmst Du Teil?**“, wurde u. a. mit den folgenden Antworten geantwortet:

- „Ich möchte ein aktives Mitglied der Tech-Community sein.“
- „Der Austausch mit anderen macht immer viel Spaß. Von jedem Workshop habe ich viel mitgenommen und interessante Frauen aus der MINT Branche kennengelernt.“
- „Wissen und Nutzen über das Thema Netzwerk erweitern.“
- „Mein LinkedIn fit machen für Recruiter“
- „Neue Menschen kennenlernen, nützliche Soft-Skills erlernen.“
- „Ich nehme an der Eventreihe STEM Girls teil, weil es mir wichtig ist, mich mit anderen jungen Frauen zu vernetzen und mehr über interessante Themen sowie Skills erfahren.“
- „Mich mit Gleichgesinnten vernetzen, spannende Persönlichkeiten kennenlernen, Blickwinkel erweitern“
- „Um mir die Bewerbungsphase und den Berufseinstieg zu erleichtern“
- „Ich freue mich auf den gemeinsamen Austausch und das Kennenlernen von anderen Frauen im IT-Bereich.“
- „Ich finde das Thema wichtig, wusste aber nicht, wie ich es umsetzen kann.“
- „Um Tipps zu bekommen, wie man seine Komfortzone verlässt und an sich wächst“
- „Weil ich oft noch zu viel Angst davor hab, meine Komfortzone zu verlassen“

Auf die Frage „**Was hat dir am besten gefallen?**“, wurde u. a. mit den folgenden Antworten geantwortet:

- „Der offene Austausch über Probleme“
- „Das lockere Setting“

Auf die Frage „**Was hast Du dazu gelernt, dass du davor noch nicht wusstest?**“ wurde u. a. mit den folgenden Antworten geantwortet:

- „Es kommt bei Bewerbungsgesprächen sehr auf den Charakter der Person an.“
- „dass es einen besseren Eindruck hinterlässt, locker an Bewerbungsgespräche heran zugehen, da man dadurch natürlicher wirkt“
- „Personaler bevorzugen Lernfähigkeit gegenüber Talent“

Auf die Frage „**Möchtest du zum Schluss noch etwas mitteilen, was noch nicht abgefragt wurde?**“ wurde u. a. mit den folgenden Antworten geantwortet:

- „Es war in allen Workshops eine wunderbar freundliche und produktive und ermutigende Atmosphäre. Ich habe mich sehr wohl gefühlt und tolle Frauen kennengelernt. Das hat mir sehr geholfen.“
- „Super, vielen Dank! Dass es kostenlos war und die Anmeldung unkompliziert, habe ich sehr genossen.“
- „Vielen Dank für die Organisation und das tolle Programm.“
- „Allen Kommilitoninnen macht das Thema Berufseinstieg viel Angst. Mir auch. Vielleicht wäre das ein gutes Thema.“
- „Ich fand den Workshop, bei dem verschiedene Frauen ihren Job vorgestellt haben echt toll und würde mich freuen wieder daran teilzunehmen“
- „Ich denke, dass es für junge Frauen am Anfang des Studiums oder Berufslaufbahn sehr hilfreich ist, sich über die angebotenen Themen auszutauschen und finde es super klasse und wichtig, dass die Workshops angeboten werden.“
- „Danke, dass es solche Workshops gibt!!“

B Ausschreibung der Events

STEM GIrls: Netzwerken 18.01.2022, 18:00 - 20:00 Uhr

Oft wirst Du in Deiner Karriere feststellen, dass ein gutes Netzwerk Dir zu dem letzten benötigten Schub verhelfen kann. Ein solches Netzwerk benötigt jedoch auch einiges an Arbeit und Pflege. Dieser Workshop wird euch ein paar Tipps und Tricks an die Hand geben, wie auch Du Dein Netzwerk verbessern und erweitern kannst.

Carolin Neumann studiert Informatik im Master an der TU Berlin, mit Fokus auf Blockchain-Technologien. Während ihres Studiums hat sie viele Praxiserfahrungen als Werkstudentin und als selbstständige Softwareentwicklerin und Data Scientist gesammelt, unter anderem in London, Luxemburg und natürlich Berlin. Ihre Faszination für Technik gibt sie nun an die nächste Generation Schüler:innen weiter. Sie ist Initiatorin und Sprecherin der Hochschulgruppe Berlin-Brandenburg der Gesellschaft für Informatik e.V. und hat die „BYTE Challenge“ im Mai 2020 ins Leben gerufen. Die BYTE Challenge ist ein 100% ehrenamtliches gemeinnütziges Projekt das komplett digital über 100 modularisierte Kurse zu MINT, Digitalen Skills und Orientierung nach der Schule deutschlandweit anbietet. Außerdem ist sie Initiatorin der Projektwerkstatt „eduhacktory“ der TU Berlin, einer Lehrveranstaltung, die digitale Lehre deutschlandweit und lokal in Berlin-Brandenburg kritisch analysiert und innovative Lösungsansätze entwickelt. Mit ihrem Engagement möchte sie die Digitalisierung voranbringen, Vorurteile über MINT-Berufe verringern und junge Frauen empowern.

Ulrike WINzer ist gefragte Beraterin, Moderatorin, Rednerin und Autorin in den Themen digitaler Wandel, Change und HR. Mit ihrer im Jahr 2008 gegründeten eigenen Beratungsfirma begleitet sie Mitarbeitende und Manager in der beruflichen Veränderung und unterstützt Unternehmen beim Wandel in die digitale Arbeitswelt. Ihr eigener Werdegang ist von Veränderung, Wandel und Digitalthemen geprägt: Softwareentwicklern, Bereichsmanagerin IT - Controlling und Einkauf, Recruiting Managerin Germany, Personal- und Unternehmensberaterin. Mit mehr als 25 Jahren Erfahrung aus Großkonzern, Mittelstand und Kleinunternehmen und der persönlichen Leidenschaft für Menschen und für Digitales setzt sie heute wirkungsvolle Impulse, um die hohe Relevanz von Veränderung deutlich zu machen und Menschen ins Handeln zu bringen. Denn die aktuellen Herausforderungen erfordern vor allem die Transformation der persönlichen Haltung.

STEM GIrls: Digital You – Souverän und erfolgreich Netzwerken 25.01.2022, 18:00 - 20:00 Uhr

Den Wald vor lauter Bäumen sehen und sich bloß nicht blamieren – das wünschen sich viele User:innen, die sich ein LinkedIn-Profil aufbauen und dieses zum zielgerichteten Netzwerken einsetzen. Berufseinsteiger:innen ebenso wie Führungspersönlichkeiten können mit einer guten Positionierung bei LinkedIn nach innen und außen Strahlkraft aufbauen, ohne dass ihnen beim Social Media Marathon die Puste ausgeht. Digital Coach Kathrin Koehler gibt Impulse für die souveräne Präsenz, einen Überblick zum themenzentrierten Netzwerken sowie aktuelle Hacks zum LinkedIn- Algorithmus.

Kathrin Koehler ist Digital Coach, Journalistin und Beraterin in den Bereichen Business Netzwerken und Remote Leadership. Sie ist überzeugt davon, dass Menschen im Berufsleben die digitale Sphäre effizient und souverän für sich nutzen können, solange die Balance zwischen Tool-Kenntnis, Strategie und Haltung stimmt. Mit der Devise „Besseres Business mit LinkedIn“ stellen ihre Digital You-Coachings das themenzentrierte Netzwerken in den Mittelpunkt und reichen methodisch von Selbstlernangeboten für den Aufbau des Profils bis zu kollaborativen Bootcamps auf C-Level – diese auch für Twitter. Mit Führungspersönlichkeiten aus Marketing und Vertrieb sowie deren Teams lotet Kathrin Koehler bereits seit 2012 die Chancen und Grenzen des Content Marketings und des Social Selling innerhalb der Business Netzwerke aus. Zu ihren Kunden zählen u.a. LinkedIn (als Rock Your Profile-Trainerin), Otto Group, Vattenfall ebenso wie Stiftungen, private und öffentliche Bildungsträger sowie europäische Sektionen amerikanischer Konzerne.

STEM GIrls: Selbstbewusst auftreten 15.02.2022, 18:00 - 20:00 Uhr

Manchmal ist es gar nicht so leicht, selbstbewusst zu sein. In dieser Session bekommst du einfache Tipps, die dir helfen, auch in herausfordernden Situationen sicher und gelassen aufzutreten. In kleinen praktischen Übungen kannst du selbst ausprobieren, welche Elemente der Körpersprache dich dabei unterstützen und wie du deine Stärken ankern kannst. Referentin Lena Johanna Schmitt ist zertifizierte Trainerin (IHK) und Business Coach (IHK). Als Senior Change Strategist bei RCKT arbeitet sie vor allem in den Bereichen digitale Kultur, Führung und Innovationen. Ihre besondere Expertise liegt in der Entwicklung und Umsetzung von individuellen Konzepten für Personal- und Organisationsentwicklung, Startups, Führung, Kommunikation und Change-Programme. Zu ihren Kunden zählen unter anderem die Commerzbank mit dem preisgekrönten Digitalisierungsprogramm DLC für Führungskräfte, die Deutsche Bahn mit verschiedenen Accelerator-Programmen und das Bundeswirtschaftsministerium mit der Digital Hub Initiative. Lena Johanna Schmitt ist seit 2008 bei der Unternehmensberatung Schmitt und Partner als Trainerin und Business Coach für Konzerne, KMUs, Startups sowie öffentliche Institutionen tätig. Seit mehr als 11 Jahren leitet sie als Dozentin für die IHK Würzburg-Schweinfurt die Zertifikatslehrgänge "Train the Trainer“ und "Business Coach". Von 2010 bis 2017 war sie bei DaWanda für den Video-Content

verantwortlich und produzierte in dieser Zeit mit ihrem Team über 500 Videos. Im Anschluss baute sie den Bereich People Innovations für das Berliner Startup auf und brachte ihre Expertise auch in den Bereichen Trainings Workshops sowie Business Coaching bei DaWanda ein. Darüber hinaus ist sie Mitbegründerin der Venture Ladies, einem interdisziplinären Frauennetzwerk mit über 4000 Mitgliedern, und engagiert sich auch pro bono in Bereichen Networking, Female Leadership und Diversity.

STEM GIrls: Komfortzone verlassen 23.02.2022, 18:00 - 20:00 Uhr

Je mehr die Welt im Außen verrückt spielt, umso mehr wünschen wir uns Kontrolle über unser Leben. Wir fühlen uns wohl im Bekannten, Verlässlichen, Absehbaren. Unkontrollierbare Veränderungen verursachen Unsicherheit. Wir sind nicht sicher, ob wir dieser neuen Situation gewachsen sind, ob wir mit den unbekannten Problemen und Aufgaben umgehen können, ob wir alle notwendigen Skills lernen können. Deshalb wagen wir viel zu selten neues, setzen zu kleine Ziele und bleiben in unserer Komfortzone. Das ist ok, wenn du nicht weiter lernen und wachsen willst. Die Historie lehrt uns aber, dass große Errungenschaften und Erfolge nicht in der Komfortzone entstehen. Außergewöhnliche Menschen, die großes erreicht haben, hatten zwar Angst, aber sind Schritt für Schritt in neue Welten eingetaucht. Sie machten Fehler, sind gefallen und wieder aufgestanden – und dabei jedes Mal ein Stück gewachsen. Wir alle tragen einzigartige Stärken in uns, die darauf warten, entdeckt zu werden. Gib ihnen die Chance, dich glücklich zu machen. Wann hast du zum letzten Mal etwas zum ersten Mal gemacht? Referentin Sabrina von Nessen ist Vorstandsmitglied eines IT-Unternehmens und nebenberufliche Speakerin, Mentorin und Autorin. Ihr Buch „Female Empowerment – Women in Tech“ ist ein kompetenter Wegweiser für weibliche Karrierestarter, angehende Führungskräfte und Unternehmerinnen, die zuversichtlich ihren Erfolg vorantreiben und ihre Ziele erreichen wollen.

STEM GIrls: Stärken finden 07.03.2022, 18:00 - 20:00 Uhr

Was sind meine Stärken? Und was möchte ich im Leben? Mit diesen wichtigen Fragen beschäftigen wir uns beim Workshop mit Katrin Bringmann. Katrin Bringmann war schon als Rednerin bei verschiedensten DAX Konzernen unterwegs und bringt uns ein super spannendes Thema mit: „Stärken finden, Ziele setzten“. Das Problem, nicht zu wissen was wir wollen, kennen wir alle. Katrin wird uns zeigen, wie man realistische Ziele für sich selber definiert und erreicht, und dafür bewusst die eigenen Stärken aktiviert. Wie du dein Leben mit diesen Skills auf der Erfolgsüberholspur lenken kannst, erfährst du am 07.03.2022 ab 18.00 Uhr!

STEM GIrls: Freelancing das digitale Nomadentum 05.04.2022, 18:00 - 20:00 Uhr

Arbeiten am Strand, mit den Füßen im Meer und Laptop auf dem Schoß - seit Covid- 19 ist der Trend vom digitalen Nomadentum noch stärker angewachsen. Für Xenia ist dieser Lebensstil seit über 2,5 Jahren Alltag. Angefangen als Vollzeit-Selbstständige, heute in Festanstellung und nebenberuflich selbstständig reist sie, ausgestattet mit Koffer und Laptop-Rucksack, wohin das Leben sie trägt. Wieso sie diesen Alltag nicht mehr missen möchte, worin aber auch die Schattenseite dieses Lebensstils bestehen, erfahrt ihr bei einem gemütlichen Beisammensein. Xenia Handel arbeitet seit über 8 Jahren im digitalen Produktmanagement. In Startups, größeren Unternehmen und in der Selbstständigkeit hat sie Erfahrungen gesammelt, die sie gerne über die Plattform CoffeeCodeBreak weitergibt, um mehr Frauen für die Tech-Welt zu begeistern. Bei Wikando baut sie als Produktmanagerin zusammen mit ihrem Team Software für gemeinnützige Unternehmen, als nebenberuflich Selbstständige unterstützt sie gerne Startups dabei, ihre digitalen Produkte zu optimieren. Kraft schöpft sie insbesondere beim Yoga oder in der Meditation, als digitale Nomadin ist sie am liebsten am Meer und beobachtet die Sonne beim Auf- und Untergehen.

STEM GIrls: IT Jobs 26.04.2022, 18:00 - 20:00 Uhr

Die IT Jobwelt ist sehr vielfältig! In diesem Event geben wir einen kleinen Einblick von drei Frauen, die von ihrem Werdegang und ihrer Arbeit aus dem Cloud Computing, Consulting und Software Engineering erzählen. Wenn Du unsicher bist, welche IT Bereiche für Dich interessant sind, ist dieses Event das Richtige für Dich! Beschreibung Drei engagierte und besondere Frauen erzählen von ihrem Werdegang und ihrer Arbeit aus dem Cloud Computing, Consulting und Software Engineering. Alle sie kommen aus verschiedenen Ländern und haben einen sehr unterschiedlichen (nicht) Informatik-Background. Surbhi gibt Einblicke in ihre Tätigkeit als Solution Engineer bei VMware. Sophia ist Beraterin bei Netlight mit dem Fokus auf Platform Engineering und Backend-Entwicklung und verrät die Vor- und Nachteile. Katya hat damals als Journalistin begonnen, ist heute aber Software Engineer bei Google.

STEM GIrls: Gehaltsverhandlungen meistern 03.05.2022, 18:00 - 20:00 Uhr

Der Weg zum richtigen Einstiegsjob kann ein langer und steiniger sein. Mit der Wunschfirma in Reichweite folgt das Gehaltsgespräch. Wie argumentierst Du richtig und was solltest Du vermeiden? Ist mehr Gehalt tatsächlich ausschlaggebend für einen besseren Jobeinstieg? Mit Maxie Duda, IT-Rekruterin bei Evolution, unterhalten wir uns über finanzielle Fragen, die sonst eher verschwiegen werden. Referentin Maxie Duda ist seit 5 Jahren erfolgreich in der IT-Personalvermittlung tätig. Eines der wichtigen Punkte ihrer täglichen Arbeit ist es Jobangebote zu managen, Kunden und Bewerbende zu verstehen, zueinander zu bringen und Lösungen zu initiieren. Keiner weiß so gut wie

Sie, welche Erwartungshaltungen beide Parteien haben und welche Herausforderungen ein Jobangebot/Gehaltsverhandlung mit sich bringen kann.

STEM GIrls: Bewerbungen rocken 23.05.2022, 18:00 - 20:00 Uhr

Frauen verhalten sich in Bewerbungsgesprächen tendenziell schüchtern und etwas zurückhaltend. Selbstvermarktung kann und muss gelernt sein. Nach diesem Workshop ist jedes Bewerbungsgespräch ein Kinderspiel! Referentin Ulrike WINzer ist gefragte Beraterin, Moderatorin, Rednerin und Autorin in den Themen digitaler Wandel, Change und HR. Mit ihrer im Jahr 2008 gegründeten eigenen Beratungsfirma begleitet sie Mitarbeitende und Manager in der beruflichen Veränderung und unterstützt Unternehmen beim Wandel in die digitale Arbeitswelt. Ihr eigener Werdegang ist von Veränderung, Wandel und Digitalthemen geprägt: Softwareentwicklern, Bereichsmanagerin IT - Controlling und Einkauf, Recruiting Managerin Germany, Personal- und Unternehmensberaterin. Mit mehr als 25 Jahren Erfahrung aus Großkonzern, Mittelstand und Kleinunternehmen und der persönlichen Leidenschaft für Menschen und für Digitales setzt sie heute wirkungsvolle Impulse, um die hohe Relevanz von Veränderung deutlich zu machen und Menschen ins Handeln zu bringen. Denn die aktuellen Herausforderungen erfordern vor allem die Transformation der persönlichen Haltung.

Making Remote Work for Women in Tech

Nicola Marsden[1] Karen Holtzblatt[2]

Abstract: Remote working during the pandemic goes far beyond the global remote work that occurred in prior years. In this paper, we look at remote working from the point of view of creating the experiences women need to thrive. Within the context of our research on retaining women in tech, we have studied the impact of remote work on women. We reveal our findings and indicate how to improve the experience of tech women in remote work.

Keywords: gender; remote work; retention

1 The @Work Experience Framework

Increasing the number of women in the tech industry is a high priority. Yet there is another issue: Once hired, women are less likely to stay in the tech industry than men [14a; 16; 20; AME16]. The first 12 years in technology seem to be the most vulnerable: 50% of women leave the field for other occupations, compared to 20% of professional women leaving non-STEM fields [Gl13]. For the European Union, it has been shown that by age 45, more than 90% of the women who graduated with a degree in information and communication technology have left the field [13].

Our research indicates that 45% of women in tech are thinking of leaving their job [HM22]. We have identified what women need to thrive and key intervention techniques to help improve the daily work practices of diverse teams. We developed the @Work Experience Framework to outline the experiences that matter to retain women in tech. The factors of the framework are: A Dynamic, Valuing Team That's Up to Something Big; Stimulating Work; The Push and Support; Local Role Models; Nonjudgmental Flexibility for Family Commitments, and Personal Power.

2 Remote Working and Retention of Women in Tech

During the pandemic, the whole tech world became a living lab with people working remotely. To understand if working remotely is good for women and if our recommendations

[1] Heilbronn University, Social Informatics, Max-Planck-Str. 39, 74081 Heilbronn, Germany https://orcid.org/0000-0003-1316-8147 nicola.marsden@hs-heilbronn.de
[2] InContext Design, Silver Springs, Maryland, USA, karen@incontextdesign.com karen@incontextdesign.com

doi:10.18420/inf2022_14

and perspectives still hold in a remote work context, we launched The Remote Work Project, doing in-depth interviews with women in tech in a variety of job roles [HM22]. The "glasses" of the @Work Experience Framework help us anticipate and be on the watch for issues. Our current findings from The Remote Work Project and other research [14b; Bi21; KPA21] revealed issues that will impact women. Here are the points—as presented in our book—that we identified as being impacted by remote work.

Creating connections. People have an increased difficulty getting to know each other both socially and as work partners. The lack of easy connection impacts team cohesion, building new relationships with work partners and managers, and enjoying the natural chitchat that allows for breaks in a co-located workplace. Remote working makes it hard for new hires to make relationships and develop a sense of belonging. The difficulty of creating and maintaining connections for all workers has a direct impact on forming Dynamic Valuing Teams.

Getting help and coaching. People, and especially new hires, feel that it is more difficult to get one-on-one help from managers and co-workers. Scheduling one-on-one meetings feels more like an intrusion when remote. New hires don't want to impose given how busy everyone is. Asking for time when it involves setting up a meeting is harder. In general, drop-in questions and help are not easy when remote. This finding tells us that working with Local Role Models and getting a Push and Support may be harder when remote.

Stimulating work. The above two findings raise the worry of whether women working remotely will get the stimulating work they want. If women receive less coaching, will they advance their skills? If they do not feel supported because it is harder to make relationships all around, will they be less likely to put themselves forward for challenging work? In a remote context, will women's contributions be less apparent to managers and co-workers? With less daily interaction will our bias against women's skills in tech disadvantage women even more?

Communicating informally. Informal communication outside of meetings suffers when everyone is remote. Side conversations, which used to happen before and after meetings and in the hallway, have nearly evaporated. Also gone are the informal sharing of work and helping each other when stuck that were common practices when co-located. Our interviewees were surprised by how much collaboration depended on dropping in. They did not realize that informal interactions drove forming shared understandings of what to do or decide and reported both errors and time wasted by formal requests through systems. Setting up meetings to chat does not substitute for dropping in and seems like an imposition. More importantly, this informal collaboration is invisible; people don't think about how to replace something when they don't know it is happening. The work, the relationships, and the sense of collaborating on something that really matters all suffer. Remote working makes it is harder for teams to collaborate dynamically, which impacts being a Dynamic, Valuing Team.

Structuring practices. We find that managers and teams have introduced more structure to their remote group working meetings. When remote, less structure and more freeform interactions in group working meetings have led to the experience of time being wasted, confusion in managing participation, and the work not getting done. Senior people have naturally gravitated to increasingly explicit practices and using remote collaborative tools to make the purpose and interactions in the meetings clear. Indeed, research finds that more structure in remote collaboration heightens the collective intelligence of the team [To21]. And more structure is good for women because the implicit becomes explicit, interactions are potentially better managed, and practices are clear. As we return to more face-to-face interactions, we hope that managers and teams will bring this increased structure with them.

Supporting life flexibility. People have emphasized that during the pandemic their manager and teams have been more willing to flex when home needs appear. Teams seem to be more understanding of home/work commitments when remote. The pandemic highlighted how the crush of home and family demands can overwhelm—and that working at home can reduce the burden because home chores and childcare can be interleaved with working. But as numerous others have pointed out, during the pandemic the burden of children and the home has fallen within gender norms; too many women are struggling to do the work and deal with home life [21a; GWY21]. Until all schools, daycare, and elder care are safe and back in place, we will not know if Nonjudgmental Flexibility will continue and how the balance of responsibility will shift as we move to more in-person work. This will have to be watched.

Increasing self-confidence. Given our overall findings about remote work, Personal Power and self-confidence will be affected. Without connection, coaching, help, and informal information, the opportunity for misunderstanding and accomplishments being overlooked is great. But we found that some managers, teams, and companies are deliberately trying to be more valuing. Self-confidence is also related to feeling psychologically safe. Interestingly, some women report feeling safer to express themselves behind the screen, more willing to participate or take a risk. And they report using side chats in a meeting to check in with each other and send notes of support if interpersonal issues get out of control [Sa21]. Of course, this works better when the camera is on, and everyone can see each other's nonverbal responses. Overall, self-confidence is clearly impacted in remote settings.

Keeping cameras on. We know from years of research on remote meetings that video on sends more cues to guide interactions than video off. But it's complicated. Most people are more comfortable with their own camera off. At the same time, people often would like to see the others in their meetings. Negotiating this contradiction is one of the challenges of remote meetings. And there is no right or wrong. There are, however, considerations regarding women that play a role here [Dh21]. We have some indication that keeping the video on in a remote setting may help women feel more able to speak up. At the same time, being on camera puts pressure on women to look good that men do not have in the same way [21b] and having video on is more fatiguing to women than to men [Sh21]. But more of our findings surround what happens with cameras off. Camera off choices can

be policies designed to respect individual choices. In meetings, some people are on video while others are not, a personal choice. But we also see people following the lead of their managers. We find that if people are new and don't know co-workers it is much harder to get to know them with the camera off [Ro20]. And some of the women we talked to report that getting negative feedback remotely with the camera off can come across harsher. From an interpersonal dynamics perspective, the camera on vs. off clearly changes the experience of working meetings. As we become more permanently remote, we will have to examine if there is a gender effect and the real impact on team cohesion and connection.

3 Conclusion

Stepping back, our research on remote experiences only makes attention to the issues raised in The @Work Experience Framework more important. A good team onboarding process is even more important when new hires and teams are remote. If all new hires, including women, do not get a good start, ensuring their success when remote will be even harder. Since more can be misperceived when remote, increasing structure when giving feedback in the processes used in working meetings is essential. Remote work needs even more explicit practices, rules of engagement, procedures to manage participation, and good facilitation. Maybe then remote work could become what one might call a leveler between men and women that some people are hoping it could be [Jo22]. Because interpersonal connection is more difficult, remote work requires even more attention to interpersonal dynamics. Increasing valuing behaviors and managing devaluing behaviors is a must in remote interactions. As for jerk behavior, we do not find more or less of it when remote. It seems that people who behave non-professionally when face-to-face also engage in these behaviors when remote, but it is more apparent to others, who may intervene. In other words, with remote work the daily work life changes, but the challenges for retaining women do not. Whether we are working face-to-face or remotely women need the same things to thrive. To make that happen, we need to deliberately design interventions ensure that remote works for women.

References

[13] Web Page, 2013, URL: http://bcwt.bg/wp-content/uploads/documents/womenActiveInICT.pdf.

[14a] Web Page, 2014, URL: https://www.ncwit.org/resources/ncwit-scorecard-report-status-women-information-technology.

[14b] Newspaper Article, 2020, April 14, URL: https://www.nytimes.com/2020/04/14/us/zoom-meetings-gender.html.

[16] Web Page, 2016, URL: https://www.ncwit.org/sites/default/files/resources/ncwit_women-in-it_2016-full-report_final-web06012016.pdf.

[20] Web Page, 2020, URL: https://www.dritter-gleichstellungsbericht.de/de/article/241.geschlechtsspezifische-arbeitskr%C3%A4ftefluktuation.html.

[21a] Web Page, 2021, URL: https://www.kompetenzz.de/aktivitaeten/arbeiten2020%20-%20Teil2.

[21b] Web Page, 2021, URL: https://www.fastcompany.com/90577160/women-dont-turn-your-camera-off-for-zoom-calls.

[AME16] Ashcraft, C.; McLain, B.; Eger, E.: Women in tech: The facts. National Center for Women & Technology (NCWIT) Colorado, CO, USA, 2016.

[Bi21] Bilotta, I.; Cheng, S. K.; Ng, L. C.; Corrington, A. R.; Watson, I.; Paoletti, J.; Hebl, M. R.; King, E. B.: Remote communication amid the coronavirus pandemic: Optimizing interpersonal dynamics and team performance. Industrial and Organizational Psychology 14/1-2, S. 36–40, 2021, ISSN: 1754-9426.

[Dh21] Dhawan, N.; Carnes, M.; Byars-Winston, A.; Duma, N.: Videoconferencing Etiquette: Promoting Gender Equity During Virtual Meetings. Journal of Women's Health 30/4, S. 460–465, 2021, ISSN: 1540-9996.

[Gl13] Glass, J. L.; Sassler, S.; Levitte, Y.; Michelmore, K. M.: What's so special about STEM? A comparison of women's retention in STEM and professional occupations. Social Forces 92/2, S. 723–756, 2013, ISSN: 1534-7605.

[GWY21] Giurge, L. M.; Whillans, A. V.; Yemiscigil, A.: A multicountry perspective on gender differences in time use during COVID-19. Proceedings of the National Academy of Sciences 118/12, S. 1–7, 2021, ISSN: 0027-8424.

[HM22] Holtzblatt, K.; Marsden, N.: Retaining Women in Tech - Shifting the Paradigm. Morgan Claypool, 2022.

[Jo22] Johnson, S.: Women deserve better: A discussion on COVID-19 and the gendered organization in the new economy. Gender, Work & Organization 29/2, S. 639–649, 2022, ISSN: 0968-6673.

[KPA21] Karl, K. A.; Peluchette, J. V.; Aghakhani, N.: Virtual Work Meetings During the COVID-19 Pandemic: The Good, Bad, and Ugly. Small Group Research/, S. 1–23, 2021, ISSN: 1046-4964.

[Ro20] Rodeghero, P.; Zimmermann, T.; Houck, B.; Ford, D.: Please Turn Your Cameras On: Remote Onboarding of Software Developers during a Pandemic. IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)/, S. 1–10, 2020.

[Sa21] Sarkar, A.; Rintel, S.; Borowiec, D.; Bergmann, R.; Gillett, S.; Bragg, D.; Baym, N.; Sellen, A.: The promise and peril of parallel chat in video meetings for work. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems/, S. 1–8, 2021.

[Sh21] Shockley, K. M.; Gabriel, A. S.; Robertson, D.; Rosen, C. C.; Chawla, N.; Ganster, M. L.; Ezerins, M. E.: The Fatiguing Effects of Camera Use in Virtual Meetings: A Within-Person Field Experiment. Journal of Applied Psychology 106/8, S. 1137–1155, 2021.

[To21] Tomprou, M.; Kim, Y. J.; Chikersal, P.; Woolley, A. W.; Dabbish, L. A.: Speaking out of turn: How video conferencing reduces vocal synchrony and collective intelligence. PloS one 16/3, S. 1–14, 2021, ISSN: 1932-6203.

3. Interdisciplinary Privacy & Security at Large Workshop (Privacy&Security@Large)

3. Interdisciplinary Privacy & Security at Large Workshop

Annika Selzer[1], Ingo J. Timm[2]

Cybersicherheit hat in den vergangenen Jahren in Politik, Wirtschaft und Öffentlichkeit enorm an Bedeutung und Beachtung gewonnen. Durch die Diffusion von Informations- und Kommunikationstechnologie in den Alltag und durch die tägliche Berichterstattung zu Schwachstellen und Sicherheitsvorfällen sind Politikern, Unternehmen, Bürgern und Verbrauchern Risiken und Verletzlichkeit von Betriebsgeheimnissen stärker bewusst geworden. Zusätzlich geben datenschutzrechtliche Vorschriften, wie die Datenschutz-Grundverordnung, Rahmenvorgaben für die Verarbeitung sogenannter personenbezogener Daten und somit für die Bedürfnisse der Bürger nach Privatheit.

Sicherheit und Datenschutz rücken daher immer mehr in den Blickpunkt vieler Akteure, die vor die Herausforderung gestellt werden, einerseits die Chancen, die durch neue Technologien entstehen, nicht gänzlich durch normative Vorhaben „auszubremsen" sowie andererseits für eine angemessene Umsetzung sicherheitstechnischer und datenschutzrechtlicher Vorgaben zu sorgen, u.a. um Betriebsgeheimnisse und das Bedürfnis natürlicher Personen nach Privatheit zu schützen. Insofern bedarf die angemessene Umsetzung von Sicherheits- und Datenschutzvorgaben unter Berücksichtigung von Chancen und Risiken neuer Technologien einer gesamtheitlichen Betrachtung aus Sicht der Informatik-, Rechts- und Wirtschaftswissenschaften.

Der Workshop „3. Interdisciplinary Privacy & Security at Large" widmet sich diesem Themengebiet. Der geplante Workshop ist interdisziplinär angelegt. Neben technischen Aspekten sollen auch rechtliche, wirtschaftliche und wirtschaftsinformatische Aspekte des Themengebiets diskutiert werden.

Für die Vorschläge des geladenen Vortrags und die Begutachtung der eingereichten Beiträge danken wir Mitgliedern des Programmkomitees Dr. Marietjie Botes (Universität Luxemburg), Dr. Matthias Enzmann (Fraunhofer SIT, ATHENE), Prof. Dr. Dennis-Kenji Kipker (Universität Bremen, IGMR), Dr. Arianna Rossi (Universität Luxemburg), Prof. Dr. Martin Steinebach (TU Darmstadt, Fraunhofer SIT, ATHENE), Prof. Dr. Thomas Wilmer (Hochschule Darmstadt), Bernhard C. Witt (it.sec GmbH) und Dr. Daniel Woods (Universität Innsbruck).

Teile des Workshops wurden vom Bundesministerium für Bildung und Forschung (BMBF) und vom Hessischen Ministerium für Wissenschaft und Kunst (HMWK) im Rahmen ihrer gemeinsamen Förderung für das Nationale Forschungszentrum für angewandte Cybersicherheit ATHENE sowie vom Bundesministerium für Wirtschaft und Klimaschutz im Rahmen des KI-Plattformprojekts SPELL unterstützt.

[1] Fraunhofer SIT | ATHENE, annika.selzer@sit.fraunhofer.de.
[2] Universität Trier, Professur für Wirtschaftsinformatik, itimm@uni-trier.de.

doi:10.18420/inf2022_15

Learning without Looking: Similarity Preserving Hashing and Its Potential for Machine Learning in Privacy Critical Domains

Marian Eleks[1], Jonas Rebstadt[2], Philipp Fukas[3] and Oliver Thomas[4]

Abstract: Machine Learning is frequently ranked as one of the most promising technologies in several application domains but falls short when the data necessary for training is privacy-sensitive and can thus not be used. We address this problem by extending the field of Privacy Aware Machine Learning with the application of Similarity Preserving Hashing algorithms to the task of data anonymization in a Design Science Research approach. In this endeavor, novel anonymization algorithms made to enable Machine Learning on anonymized data are designed, implemented, and evaluated. Throughout the Design Science Research process, we present a collection of issues and requirements for Privacy Aware Machine Learning algorithms along with three Similarity Preserving Hashing-based algorithms to fulfil them. A metric-based comparison of established and novel algorithms as well as new arising opportunities for Machine Learning on sensitive data are also added to the current knowledge base of Information Systems research.

Keywords: Data Anonymization, Similarity Preserving Hashing, Machine Learning, Fuzzy Hashing, Privacy Aware Machine Learning

1 Introduction

While Machine Learning (ML) is utilized in many application domains like healthcare [WHZ13], auditing [Fu21], education [Sc22] or smart living [Ba19], [Ko22], it is often hindered by privacy-sensitive training data. Often, ML-projects fail in practice because the required training data cannot be used [Fi20]. What is needed to overcome these challenges is a type of ML that ensures the privacy of sensitive training data through various actions, also named "Privacy Aware Machine Learning" (PAML). Techniques that suit PAML have been explored since the late 1990s [Sw97] and early 2000s [Sw02b] first for the purpose of data sharing and later specifically for ML [Ma16]. Nowadays, a plethora of techniques have been developed or applied but every one of them comes with its own

[1] Strategion GmbH, Albert-Einstein-Straße 1, 49076 Osnabrück, marian.eleks@strategion.de, https://orcid.org/0000-0002-1516-5129
[2] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, jonas.rebstadt@dfki.de, https://orcid.org/0000-0001-8531-3273
[3] Universität Osnabrück, Informationsmanagement und Wirtschaftsinformatik, Parkstraße 40, 49080 Osnabrück, philipp.fukas@uni-osnabrueck.de, https://orcid.org/0000-0002-2169-5505
[4] Universität Osnabrück, Informationsmanagement und Wirtschaftsinformatik, Parkstraße 40, 49080 Osnabrück, oliver.thomas@uni-osnabrueck.de

doi:10.18420/inf2022_16

drawbacks, making them unapplicable to a number of scenarios and leaving gaps in PAML-knowledge that this work aims to fill.

To address the shortcomings of the state-of-the-art, we propose a new technique for PAML. At its core, the novel technique transfers Similarity Preserving Hashing (SPH), also known as fuzzy hashing, practices from their original domain of data forensics [GHH14] to data anonymization to ensure anonymized datasets that still contain the relevant information needed for ML. SPH at its core is a type of hashing that does not just detect identical data but also data with small changes, which means that Similarity Hashes both contain information needed to separate and compare data as well as being irreversible transformations of data, making it a promising candidate for PAML. To achieve the goal of producing SPH-based PAML algorithms, a Design Science Research (DSR) approach following Hevner et al. [He04] and Peffers et al. [Pe07] is applied. As an initial step, general shortcomings of current PAML techniques are identified through a systematic literature review based on Webster & Watson [WW02]. These are then translated into general requirements for the implementation of the SPH-based solution. To address the identified requirements, promising algorithms are designed and implemented as core artifacts of the DSR-process. The implemented solution is artificial technical evaluated using metrics to measure both its utility for ML as well as its privacy-preservation capabilities in comparison to existing techniques on an adult income dataset [RU96b]. To structure this procedure, we pose three research questions:

RQ 1: What problems arise in the application of current PAML-techniques?
RQ 2: How can SPH algorithms be adapted for use in data anonymization for PAML?
RQ 3: What level of privacy and utility can SPH anonymization methods achieve compared to established techniques identified in RQ 1?

The paper begins by introducing the general concepts of PAML and SPH along with related work, followed by explaining the methodology in more detail. Afterwards, the full DSR iteration along with its findings is described. Finally, the findings of the DSR process are discussed and a conclusion and an outlook on future research are given.

2 Privacy Aware Machine Learning with Similarity Preserving Hashing

Although the term “Privacy Aware Machine Learning” was first used by Malle et al. [Ma16] in 2016 and “Privacy Aware Learning” by Duchi, Jordan and Wainwright [DJW14] in 2014, the goal of sharing and analyzing data while protecting sensitive attributes was pursued decades earlier. Before Sweeney introduced k-Anonymity in 2002, the goal of privacy protection was achieved by suppressing attributes that directly identify individuals in the data [Sw02b]. Sweeney [Sw02b] demonstrates that there also exist quasi-identifiers besides the direct identifiers mentioned. Quasi-identifiers are a combination of attributes that can identify individuals together. She was able to identify

87% of the American people with a combination of ZIP code, date of birth and gender. Because of this, she introduced the concept of k-Anonymity: Every unique combination of quasi-identifiers must occur in at least k different times in the dataset and thus guarantee that combinations of quasi-identifiers cannot uniquely identify an individual [Sw02b]. In subsequent research, problems with the privacy aspect of k-Anonymity were found and the refining concepts of l-Diversity and t-Closeness were introduced in 2007 [LLV07], [Ma07]. Besides these techniques, the concept of differential privacy was introduced by Dwork in 2006 [Dw06]: Aggregated information about a database is published in a way that limits the disclosure of private information about an individual in said database given a "scaling factor" ε, which influences the privacy guarantee of the employed privacy mechanism [DR13], [Dw06]. Besides k-Anonymity and differential privacy, some newer research uses either homomorphic encryption or synthetic data to achieve the goals of PAML [LZ20]. Homomorphic encryption, initially introduced in 1978 by Rivest et al. [RSA78], is a type of asymmetric encryption that allows mathematical operations on the encrypted data, giving the correct result once decrypted [YPB14]. Since ML often consists of a complex string of mathematical operations, it can be done entirely on encrypted data using homomorphic encryption and thus ensuring privacy [LZ20]. Synthetic data on the other hand takes a different approach: Instead of using the original dataset containing sensitive data, techniques like generative adversarial networks (GANs) [FMT22], [Fu22], [PMJ18] are used to generate synthetic data that by its nature does not have any connection to real individuals or events and is therefore privacy-preserving. The data synthesis moreover guarantees that the general structure of the original data is kept, ensuring the use for ML. To summarize, k-Anonymity, differential privacy, homomorphic encryption, and synthetic data all pursue the goals of PAML but come with their own advantages and shortcomings. They all aim to anonymize or pseudonymize the data respectively, obscuring connections to individuals. Both the techniques themselves as well as their benefits and challenges are collected in a systematic literature review (see section 3).

Since this paper introduces a new PAML technique originating from SPH, a brief explanation of the characteristics of SPH that make it suitable for PAML-compliant anonymization is given. As the name suggests, SPH is a subcategory of Hashing using selected characteristics of cryptographic Hashing and adding new characteristics to suit their purpose based on document comparison for data forensics [GHH14]. The purpose of hashing in general is to map any input I to an output H(I) using a hash function. An input always produces the same output, but different inputs result in different outputs (*collision resistance*). Cryptographic hashing functions go even further and ensure that even slightly varying inputs produce vastly different outputs (*near-collision resistance*). They also require it to be impossible to map an output of a hash function back to its input (*image resistance*). Furthermore, it should not even be possible to generate an Input I' based on an original input I that results in the same output value (2^{nd}*-image-resistance*). These cryptographic hashing characteristics were gathered by [ADB11]. SPH adopts all these characteristics except for the near-collision-resistance since it is used not only to identify identical data but also similar data like previous versions of a document. Breitinger and Baier [BB12b] establish five additional characteristics for SPH in addition to the adopted

ones from cryptographic hashing. Firstly, since large numbers of files must be compared efficiently, SPH functions should be easy to compute (*ease of computation*) and result in hash-values that are much shorter than their input data (*compression*). Additionally, a similarity metric between hashes needs to be produced to easily identify how similar two inputs are (*similarity score*). To ensure that changes anywhere in a file are detected, every byte of an input needs to influence the output (coverage). It should also be impossible to make intentional changes to input data that keep the data similar but not result in a match (*obfuscation resistance*).

For PAML, the image resistance guarantees output data that cannot be traced back to its input and thus it ensures privacy. The compression along with ease of computation promise good performance, which, although not being a primary goal of PAML, is always advantageous. Finally, the similarity score and full data coverage result in an output that still contains information about similar and dissimilar attributes (and labels for supervised learning), promising data that is highly useful for ML.

3 Methods

This paper uses a research approach based on the Design Science Research (DSR) paradigm of Hevner et al. [He04]. DSR targets the development and evaluation of artifacts such as "constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems)" [He04]. Utilizing the methodical approach suggested by Peffers et al. [Pe07], the research process is structured based on Peffers et al. [Pe20] in six steps *(1) Identify Problem & Motivate, (2) Define Objectives of a Solution, (3) Design & Development, (4) Demonstration, (5) Evaluation and (6) Communication*.

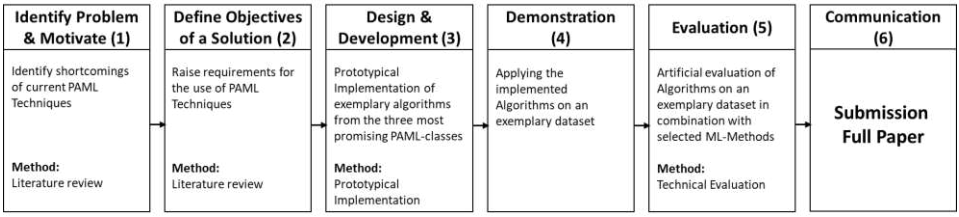


Fig. 1: Research Approach based on Peffer et al. [Pe20].

As shown in Fig. 1, the research process starts with *(1) Identify Problem & Motivate* and *(2) Define Objectives of a Solution.* These two steps are addressed using a systematic literature review as proposed by [WW02]. To identify and collect all relevant research, standardized steps of a systematic literature review are conducted [Vo09]. The review scope is defined [Co88], central articles are identified and the topics are conceptualized to specify the key terms for the search string [RS04]. Using this, the following search string is defined: *"anonymization" AND ("technique" OR "tool" OR "algorithm") AND*

"machine learning". IEEE Xplore, Springer Link, Google Scholar and *AISeL* are browsed using this search string and initially lead to 16619 search results, which are scanned based on their titles and abstracts resulting in 195 possibly relevant results. Finally, duplicates are eliminated, and the articles are analyzed towards our inclusion and exclusion criteria resulting in 52 relevant publications. Setting up on the results, three algorithms from the most promising PAML-classes based on [GHH14] are implemented in the *(3) Design & Development* phase. The implemented algorithms are then applied on an exemplary dataset in the *(4) Demonstration* and artificially evaluated in cooperation with selected ML techniques and metrics to show the resulting performance in the *(5) Evaluation*.

4 Addressing Shortcomings of Current PAML Techniques

4.1 Shortcomings of current PAML techniques

This section represents the initial phase of the DSR approach, identifying problems of current PAML anonymization methods based on the conducted systematic literature review along with their innate properties and thus answering RQ 1. In the literature review, four main anonymization techniques were identified: *(1) synthetic data, (2) k-Anonymity, (3) differential privacy*, and *(4) homomorphic encryption*. Section 2.1 introduces them in general whereas this section aims to identify problems that can later be solved by the newly designed SPH algorithms.

The core problem of *(1) synthetic data* is the application of ML-methods for data synthesis like for example autoencoders [Ab19] or generative adversarial networks [Pa18]. Therefore, synthetic data generation represents an ML instance in itself, which also requires training data for its use, again raising the mentioned problems of potentially privacy-critical training data. In addition, methods based on neural networks, such as autoencoders or GANs, require a large amount of data [Br17] and time for training, which combines data synthesis for anonymization with other practical problems. Moving to the problems of *(2) k-Anonymity*: It requires the identification of (quasi) identifiers [Sw02a]. Since these can be (subjectively) different for each dataset, identification can only be done manually, which both stands in the way of process automation and introduces the possibility for human error. This situation is not a good condition for ML on anonymized data. Additionally, with k-Anonymity, longer runtimes must be expected to achieve the highest utility while preserving the privacy requirements [Ma16], [MW04].

Other problems concern ML with data anonymized via *(3) differential privacy*: differential privacy does not output microdata, only data aggregated and perturbated with a function [Dw06], which makes it impossible to train for instance a model that is supposed to classify individual rows of data. In addition, models such as differential privacy or k-Anonymity, which publish a (processed) version of the original data in a human-readable form, always have the problem that they have to reduce either the information content of

the data through suppression or generalization [Ma16], [Sw02a] or the truth content of the data through perturbation [DR13] in order to achieve their privacy goals.

Human readability is another noteworthy problem. Transforming the data into a non-human-readable space could be a promising step and *(4) homomorphic encryption* provides such a transformation [LZ20]. However, it also brings its own problems: First of all, it has the problems inherent in all encryptions, namely the need to store a key and the human component. Should an unauthorized person get access to the key to the encrypted data, the security and anonymity of the individuals in the data is effectively nullified. One way this scenario can occur is through the human factor. People can forget keys, store them insecurely or pass them on unknowingly, compromising the privacy of homomorphic encryption. In addition to these problems, complete homomorphic encryption (all mathematical operations possible) can currently only be implemented with either cumulative noise in each operation or massive processing time [Co21]. These factors limit the many advantages of homomorphic encryption. Another step after homomorphic encryption would be to transform the attributes into a different data space so that Quasi Identifiers and Sensitive Attributes would disappear entirely. This is done for example in Principal Component Analysis (PCA), which is used in practice for pseudo-anonymization, e.g. in the Credit Card Fraud Detection dataset [Ul18]. Unfortunately, PCA – just like homomorphic encryption – is reversible, which is why it cannot be considered as an anonymization method [Am16].

4.2 Define Objectives of a Solution

After the current state of data anonymization and its problems have been determined in the last subsection, requirements for the subsequent design phase are systematically derived from the reviewed literature.

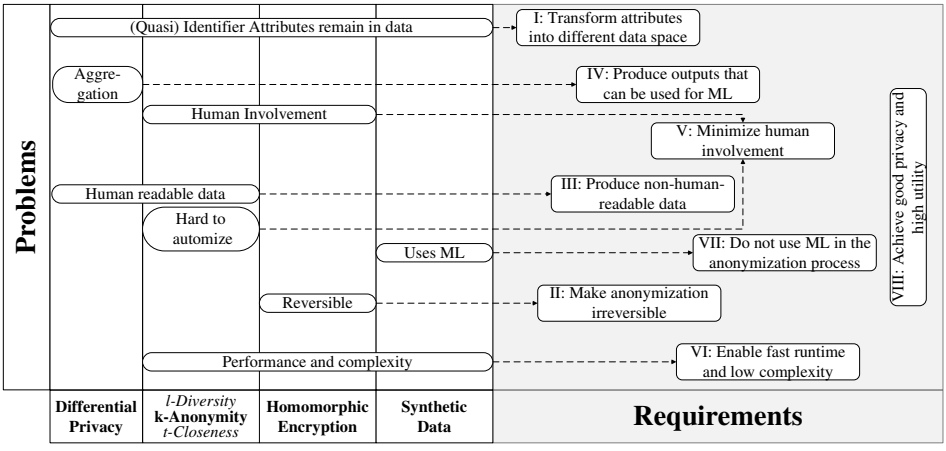


Fig. 2: Requirements of the solution

As shown in Fig. 2, there are a total of eight requirements that need to be considered when designing SPH anonymization. Output attributes should be transformed into a different data space to eliminate (quasi) identifiers. The output data should also be irreversible to the input and non-human-readable while still being suitable for ML. The anonymization process should require as little human involvement as possible. It should also be quick, with little effort, and not use ML itself. Ultimately, the goal is to develop an algorithm that achieves a high degree of privacy and utility under these requirements.

4.3 Design & Development

Having introduced current problems and resulting requirements of anonymization for ML as well as SPH in general, this subsection deals with designing and implementing SPH algorithms for anonymization and thereby answering RQ 2. Gayoso Martínez et al. [GHH14] classify SPH algorithms into four categories: Block Based Hashing (BBH), Context Triggered Piecewise Hashing (CTPH), Statistically Improbable Features (SIF) and Block-Based-Rebuilding (BBR). We will implement one algorithm of every category except BBH since "a simple character insertion/deletion towards the beginning of the file could render all block hashes different" [GHH14] making it unsuitable for anonymizing different types of inputs. All types of algorithms share some central aspects, which will be introduced beforehand: First, the data is processed in its binary representation. All algorithms choose parts of the data by their specific selection criteria to contribute to the final hash. Lastly, where SPH algorithms are usually used to compare entire files, the input in this use case are single rows of a tabular dataset.

While fitting the SPH algorithms for anonymization, some overarching changes for all SPH algorithms are made: In general, the desired output of the algorithms is not one singular hash-value, but instead several values that form the new attributes of the data. In all the original SPH-algorithms, at some point during the process, there exists a collection of values that can be used to generate the mentioned attributes. These values can either directly be used as new attributes or they are a collection of hashed selected parts of the data, which will be called "relevant hashes". Should the latter come up, attributes are generated by counting occurrences of the relevant hashes inspired by the sum table used to generate the final hash in the SimHash algorithm [SL07]. These procedures are explained in more detail in the following description of the algorithms.

Ssdeep is an example of a CTPH algorithm, identifying context markers to establish data chunks that are then hashed and processed further to generate a final hash. It was created in 2006 by Kornblum et al. [Ko06] and works by computing a rolling hash over the binary data using a specific window size and the adler32 hash function [Ad95]. The resulting hash is inserted into a semi arbitrary equation and if the equation holds, is marked as a trigger. These triggers partition the data into chunks that are then hashed using the Fowler-Noll-Vo (FNV) hash function [Fo91]. Ssdeep would originally compute a Base64 letter for every chunk, all of which together make up the final similarity hash. Comparability is enabled by the Levenshtein Distance [Le66]. The changes made for anonymization start

after the chunks are generated. Instead of using FNV-hash, SHA-256 [Na04] is employed, because FNV does not bring the guarantees of cryptographic hashing functions [Fo91]. After that, instead of generating Base64 letters, the frequency of each hash value is calculated over every row in the data. Hashes that appear in almost all the rows are filtered out since they make for bad splits during ML. Likewise, hashes are also filtered out if they appear in next to none of the rows since they also hold little information about the data in general. The remaining counted hashes make up the final anonymized attributes, structured like in the sum table constructed for SimHash [SL07].

The **sdhash** algorithm by Roussev et al. [Ro09] falls into the SIF category, identifying sequences in the binary data that are likely to occur in the observed datum but not in others, and uses entropy as its identification metric. It starts by moving over the binary data with a sliding window of a certain size. The collection of all sequences in the data (feature list) forms a starting point. Roussev et al. [Ro09] then calculate the normalized entropy for every sequence. Since entropy is the amount of surprise or uncertainty, the sequences with the lowest entropy are best at splitting the data. Thus, Roussev et al. [Ro09] calculate a popularity value k for every sequence based on how low the normalized entropy is. Another sliding window calculation is initialized over the feature list. For every slide, the feature with the lowest entropy has its popularity k raised by 1, starting at 0. Afterwards, only features with large enough k are chosen to be the input for the similarity hash. They are then hashed with a cryptographic hash function and inserted into a bloom filter. Two outputs are more similar the more features they have in common. The changes for anonymization for sdhash are minimal. Instead of inserting the hashed features into a bloom filter, they are counted row-wise to make the feature/attribute values like for ssdeep. Additionally, the probabilities used to calculate the entropy of a feature for anonymization are defined as $P(X_0) = \frac{count(S)}{n}; P(X_1) = 1 - P(X_0)$ with *count(S)* being the number of occurrences of a feature and *n* being the number of all features.

Finally, **bbHash** was created by Breitinger and Baier in 2012 [BB12a] and falls into the BBR category, searching for pre-assigned patterns in the data. Adhering to the name "block-based", it starts by generating 16 byte-sequences of length 128, the building blocks. After that, just as in the former two algorithms, a rolling window is computed over the input data and for every window, the Hamming distance to all building blocks is calculated. The block with the smallest distance contributes to the final hash if the given distance is smaller than a threshold k. This threshold is calculated so that the coverage is ensured by solving equation (1) by Breitinger and Baier [BB12a]. Breitinger and Baier [BB12a] use the index of a building block to contribute to the final hash, so the similarity hash of a file is made up of hexadecimal digits denoting the index of chosen building blocks during the rolling window computation. To compare hashes, the authors suggest either using the Hamming distance like in Roussev et al.'s [Ro09] sdhash or weighted edit distance (Levenshtein) like in Kornblum et al.'s [Ko06] ssdeep. To adapt this algorithm for data anonymization, bit sequences are used instead of byte sequences since rows of input data are usually much shorter than files. Additionally, the normalized Hamming distance is used instead of the overall Hamming distance Breitinger and Baier [BB12a]

use. This places the distance of any sequence to any building block between 0 and 1 so that a constant threshold parameter k can be used, which is chosen experimentally and set to 0.45. Finally, the indices of building blocks cannot be immediately used as attributes for ML. Different rows of input data generate different numbers of indices, so it is necessary to adjust that number by filling outputs with zeros to the length of the longest output. The padded index rows can then be used as input for ML either directly or with one-hot encoding since a larger index has no semantic meaning as a larger number of hashes would have in the previous algorithms.

4.4 Demonstration & Evaluation

To answer RQ 3, the demonstration and evaluation takes place in terms of DSR. Since at this stage the designed artifacts are purely technical, we followed an artificial evaluation strategy [VPB16]. Specifically, this means that both SPH and existing PAML techniques are used to anonymize data for ML. Afterwards, utility and privacy are measured. The dataset used in this work is the Adult Income dataset [RU96], which is further described by Revow [RU96b]. It contains 14 attributes and is labelled with a binary determining whether an individual makes more or less than $50,000 a year. Since it contains both a sensitive label and quasi-identifiers, the dataset is suitable for the purpose of this work.

The two ML methods used are an Artificial Neural Network (ANN) and a Random Forest (RF). Tensorflow is used to construct the ANN [Ab15], [Ch15] as a dense sequential model [ReLu→Dropout→ReLu→Sigmoid]. The input dimensions are adapted to the selected input data, but the model otherwise follows [input dimensions→12→12→1]. In training, a batch size of 20 is chosen and a class weighting is introduced since the Adult Income dataset has a slight imbalance. For our research, a simple ANN is sufficient since our goal is finding differences in data quality and not training the best possible model. The second model used for the evaluation is a RF. It is implemented using the package "scikit-learn" [Pe11] with 100 estimators and entropy as splitting criterion. After configuration of the ML training, all anonymization algorithms can be applied to the data. It is assumed that all attributes in the Adult Income Dataset are to be anonymized, so that a comparability of all algorithms with SPH anonymization, which always anonymizes all attributes, is given. The existing PAML algorithms found during the literature review are differential privacy, k-Anonymity, homomorphic encryption, and synthetic data (see section 4.1). From those methods, only the last three methods are suitable because differential privacy outputs aggregated data, which cannot be used to anonymize data meant for individual classification and is thus not comparable for SPH-anonymization.

The first anonymization method is **k-Anonymity**. Instead of creating a new implementation with possible errors, the Mondrian algorithm following LeFevre, DeWitt, and Ramakrishnan [LDR06] is used. It offers a good combination of anonymity and utility according to Gong et al. [GKM19]. A Python implementation of the algorithm was created by Gong et al. [GKM19] and can be applied directly to the Adult Income Dataset with the default setting of k=10. As a result of the algorithm, 4044 10-anonymous clusters are

produced from the 48842 rows, which can be further processed and used for ML in a manner analogous to the original data. Although the use of **data synthesis** is applying ML itself, it is listed as an option in the literature and therefore tested. For practical implementation the Synthetic Data Vault [Pa16], a project of the Data to AI Lab at MIT, is chosen. The Conditional GAN model presented in Xu et al. [Xu19] is chosen to generate tabular data from the original dataset. Ultimately, the trained model can be used to generate any number of rows that are structured analogous to the original data. **Homomorphic encryption** is also tested using an existing library, namely Tensorflow-Encrypted (TF-Encrypted) [Da19]. A test run from the original TF-Encrypted samples works flawlessly [DJ20], but running it on the significantly more complex Adult Income dataset yields meaningless losses between negative and positive billions, which exists because the cumulative noise has gotten out of hand. Just like the loss, the output of the trained model is also meaningless noise. In TF-Encrypted, there is no way to reduce noise, but an unjustifiable runtime is to be expected for higher noise budgets or noise-free encryption schemes [Mu20]. Due to this, homomorphic encryption cannot be compared to SPH with the chosen metrics. Instead, it serves as an example for a technology that has a lot of potential for PAML but is held back by its inherent problems at the current time.

After the implementation of the established and new anonymization methods, the evaluation is performed. Utility, measured by F_1-Score [SL09] and accuracy, and privacy, measured by mutual information [Sh48], are evaluated for the three SPH algorithms as well as the k-anonymized data, synthesized data and original data. The results are depicted in Tab. 1 for the ANN and RF. To measure privacy, a quantifiable information theoretic metric from the taxonomy of Wagner and Eckhoff [WE18], namely Mutual Information is chosen. It was introduced by Claude Shannon [Sh48] as Rate of Actual Transmission receiving the name Mutual Information in 1957 [Kr57]. Mutual Information is applied to quantify how much information can be found out about an original variable X when its anonymized version Y is known. In the case of this work, the variables are attributes in the datasets like age or gender. In the case of the SPH algorithms, the attributes change, so for each of the 14 original attributes, the SPH attribute with the highest Mutual Information is chosen as the base for measurement. The sum of Mutual Information of all attributes is used as the final metric (see Tab. 1). Less Mutual Information is seen as better Privacy, since less information about the original data is revealed.

	Accuracy (in %)		Precision		Recall		F_1-Score		Mutual Information
	ANN	RF	ANN	RF	ANN	RF	ANN	RF	
No Anonymization	85,4	85,4	0,70	0,73	0,66	0,61	0,68	0,67	42,72
k-Anonymization	80,3	81,5	0,59	0,66	0,51	0,50	0,58	0,57	11,77
synthetic data	83,0	83,7	0,63	0,69	0,67	0,58	0,65	0,63	14,57
ssdeep	83,7	82,6	0,68	0,67	0,59	0,54	0,63	0,60	10,06
sdhash	84,4	83,3	0,68	0,68	0,65	0,56	0,66	0,62	10,89
bbHash	80,4	85,4	0,57	0,67	0,58	0,37	0,59	0,48	1,52

Tab. 1: Evaluation Metrics

The combined results of the evaluation can be seen in Fig. 3. The top left corner of the diagram the area that a PAML-algorithms strives for: Lowest possible privacy loss (mutual information) with highest possible utility (F_1-Score).

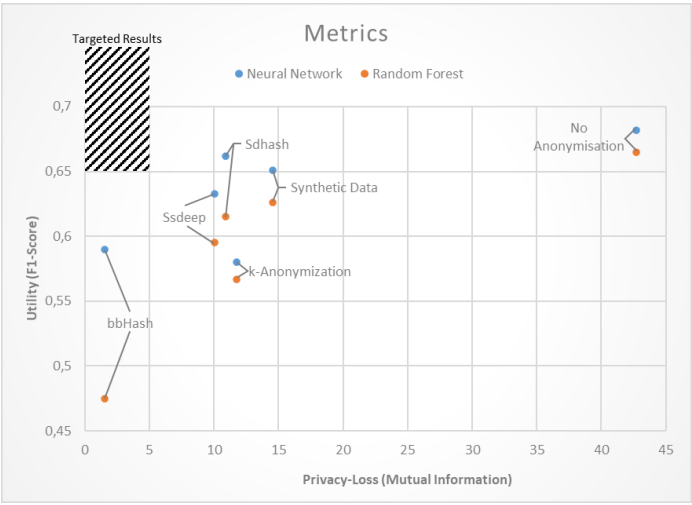


Fig. 3: Privacy and Utility Metrics

Giving a baseline comparison, not anonymizing at all entails severe privacy loss with high utility. The best privacy is achieved by bbHash and the best utility by the sdhash ANN. RFs always achieve slightly lower utility, especially for bbHash, which is due to the high number of attributes that the RF is worse at handling than the ANN.

5 Discussion

As shown in section 4.1, current PAML-techniques are subject to several shortcomings to which the use of SPH algorithms is a solution. To evaluate the usefulness of SPH algorithms for PAML, in section 4.2 eight central requirements have been derived. The first three requirements, named “Transform attributes into different data space”, “Make anonymization irreversible” and “Produce non-human-readable data” are all natively fulfilled by SPH by generating an output that meets the mentioned requirements. The produced output can directly be used for ML and therefore, the fourth requirement “Produce output that can be used for ML” and the fifth requirement, addressing the minimization of human involvement, are satisfied. The sixth requirement “fast runtime and low complexity” on the other hand, is only partially fulfilled. While sdhash and ssdeep are time-efficient, bbHash is inherently non-performant. An added value of the SPH algorithms is also the implementation of anonymization without the implied ML that

would be necessary for synthetic data. Thus, requirement seven, “Do not use ML in the anonymization process”, is also fulfilled. The last remaining requirement is key to the success of SPH algorithms, but it is also the most difficult to evaluate based on the results presented in the last sections. The required high level of privacy and utility is different for the three SPH algorithms considered. While ssdeep and sdhash outperform k-Anonymity in terms of both privacy and utility, there are clear differences for bbHash. BbHash clearly offers the best privacy measured by mutual information, but the utility deteriorates significantly when used for training RFs. However, this is due to problems of the RF with high attribute numbers. The interaction can be significantly optimized by adjusting the block length to a smaller number of attributes when anonymization with bbHash, resulting in similar utility values as with ANNs. Furthermore, the explainability of the algorithms used and the decisions made has emerged as a key factor for many use cases [Fu22], [Re22]. The lack of readability of the input values due to anonymization makes it much more difficult to provide explanations. Next to this algorithmic limitation, there are two limitations in our procedure we need to address. First, differential privacy and homomorphic encryption are not included in the evaluation due to the issues mentioned in section 4.4, even though they could provide promising results. Moreover, the evaluation of privacy was done with the help of mutual information. Even though mutual information is very well suited to measure the privacy loss, the determination of the “time to adversary’s success” in a practical penetration test crucial for the practical application of the algorithms.

Our results provide great added value from both a theoretical and a practical perspective. Based on the elicitation of requirements for PAML algorithms, insights from data forensics are applied to anonymization to fill identified gaps in the current literature. Furthermore, concrete algorithms are evaluated and compared from both utility (F_1-Score) and privacy (mutual information) perspectives. Based on these findings, new opportunities for the practical application of ML on sensitive data sources are highlighted to enable rapid practical application.

6 Conclusion and Outlooks

In conclusion, this paper has successfully navigated a full iteration of DSR to create and evaluate novel anonymization algorithms for PAML to address shortcomings of this very field. RQ1 is answered through the collection of issues in section 4 and RQ2 is answered through the entire design process depicted in section 5. Lastly, the evaluation and comparison answer RQ3, the newly developed algorithms achieve privacy and utility levels comparable or higher than current techniques and simultaneously solve the identified issues. As contribution, our work provides the algorithms themselves with the measures to fit them into an anonymization setting as well as the issues and requirements for PAML-techniques on which they are based. Additionally, we provide an evaluation of current and novel algorithms on privacy and utility metrics along with the following suggestions for future research: A penetration test for the SPH algorithms is needed before

any practical use. Also, applying the novel algorithms to different datasets with different parameters will consolidate their value for PAML. Finally, since any data can be fed into the novel algorithms as long as it can be represented in binary, various data types like images or texts can be used as input and SPH algorithms may even be able to generate features for ML from these data types in addition to their anonymization capabilities.

7 Literature

[Ab15] Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G. S.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Goodfellow, I.; Harp, A.; Irving, G.; Isard, M.; Jia, Y.; Jozefowicz, R.; Kaiser, L.; Kudlur, M.; Levenberg, J.; Mané, D.; Monga, R.; Moore, S.; Murray, D.; Olah, C.; Schuster, M.; Shlens, J.; Steiner, B.; Sutskever, I.; Talwar, K.; Tucker, P.; Vanhoucke, V.; Vasudevan, V.; Viégas, F.; Vinyals, O.; Warden, P.; Wattenberg, M.; Wicke, M.; Yu, Y.; Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. 12th $USENIX$ Symposium on Operating Systems Design and Implementation, 2015, S. 265–283.

[Ab19] Abay, N. C.; Zhou, Y.; Kantarcioglu, M.; Thuraisingham, B.; Sweeney, L.: Privacy preserving synthetic data release using deep learning. ECML PKDD 2018: Machine Learning and Knowledge Discovery in Databases, 2019, vol. 11051 LNAI, S. 510–526.

[Ad95] Adler, M.: zlib Changelog - Version 0.4, zlib, 1995. URL: https://zlib.net/ChangeLog.txt (09. August 2021).

[ADB11] Al-Kuwari, S.; Davenport, J. H.; Bradford, R. J.: Cryptographic Hash Functions: Recent Design Trends and Security Notions *. S. 1–37, 2011.

[Am16] amoeba: How to reverse PCA and reconstruct original variables from several principal components?, stackexchange, 2016. URL: https://stats.stackexchange.com/questions/229092/how-to-reverse-pca-and-reconstruct-original-variables-from-several-principal-com (06. August 2021).

[Ba19] Bauer, J.; Hoffmann, H.; Feld, T.; Runge, M.; Hinz, O.; Mayr, A.; Förster, K.; Teske, F.; Schäfer, F.; Konrad, C.; others: ForeSight-platform approach for enabling ai-based services for smart living. International Conference on Smart Homes and Health Telematics, 2019, S. 204–211.

[BB12a] Breitinger, F.; Baier, H.: A Fuzzy Hashing Approach Based on Random Sequences and Hamming Distance. ADFSL Conference on Digital Forensics, Security and Law, 2012, S. 89–101.

[BB12b] Breitinger, F.; Baier, H.: Properties of a similarity preserving hash function and their realization in sdhash. 2012 Information Security for South Africa - Proceedings of the ISSA 2012 Conference, 2012, S. 1–8.

[Br17] Brownlee, J.: How Much Training Data is Required for Machine Learning?, Machine Learning Mastery - Machine Learning Process, S. 1, 2017.

[Ch15] Chollet, F.: Keras, GitHub repository, 2015. URL: https://github.com/fchollet/keras (01. September 2021).

[Co21] Constantin, L.: Intel, Microsoft join DARPA effort to accelerate fully homomorphic encryption | CSO Online, Chief Security Officers Online, S. 1, 08-Mar-2021.

[Co88] Cooper, H. M.: Organizing knowledge syntheses: A taxonomy of literature reviews. In: Knowledge in Society 1 (1988) 1, S. 104–126.

[Da19] Dahl, M.; Patriquin, J.; Dupis, Y.; Jvmncs; DeCoste, B.; Giraud, M.; Livingstone, I.; Alibaba Gemini Lab: Introducing TF-Encrypted - Alibaba Gemini Lab, GitHub, 2019. URL: https://alibaba-gemini-lab.github.io/docs/blog/tfe/ (01. September 2021).

[DJ20] Dahl, M.; Jvmncs: TF-Encrypted, Keras-Training, GitHub. Cape Privacy, S. 1, 2020.

[DJW14] Duchi, J. C.; Jordan, M. I.; Wainwright, M. J.: Privacy aware learning. In: Journal of the ACM 61 (Nov. 2014) 6, S. 1–57.

[DR13] Dwork, C.; Roth, A.: The algorithmic foundations of differential privacy. In: Foundations and Trends in Theoretical Computer Science 9 (2013) 3–4, S. 211–487.

[Dw06] Dwork, C.: Differential Privacy. International Colloquium on Automata, Languages, and Programming, 2006, vol. 4052 LNCS, S. 1–12.

[Fi20] Fischer, L.; Ehrlinger, L.; Geist, V.; Ramler, R.; Sobieczky, F.; Zellinger, W.; Moser, B.: Applying AI in Practice: Key Challenges and Lessons Learned. International Cross-Domain Conference for Machine Learning and Knowledge Extraction, 2020, vol. 12279 LNCS, S. 451–471.

[FMT22] Fukas, P.; Menzel, L.; Thomas, O.: Augmenting Data with Generative Adversarial Networks to Improve Machine Learning-Based Fraud Detection. Wirtschaftsinformatik 2022 Proceedings, 2022.

[Fo91] Fowler, G.; Noll, L. C.; Vo, K.-P.; Google; Eastlake, D.; Huawei Technologies; Hansen, T.; AT&T Laboratories; Cisco Systems: The FNV Non-Cryptographic Hash Algorithm, Network Working Group. S. 1–119, 1991.

[Fu21] Fukas, P.; Rebstadt, J.; Remark, F.; Thomas, O.: Developing an Artificial Intelligence Maturity Model for Auditing. 29th European Conference on Information Systems Research Papers, 2021.

[Fu22] Fukas, P.; Rebstadt, J.; Menzel, L.; Thomas, O.: Towards Explainable Artificial Intelligence in Financial Fraud Detection : Using Shapley Additive Explanations to Explore Feature. International Conference on Advanced Information Systems Engineering: CAiSE2022, 2022, vol. 3, S. 109–126.

[GHH14] Gayoso Martínez, V.; Hernández Álvarez, F.; Hernández Encinas, L.: State of the Art in Similarity Preserving Hashing Functions. The 2014 International Conference on Security and Management (SAM'14), Worldcomp 2014, 2014, S. 139–145.

[GKM19] Gong, Q.; Kun, L.; Mohanty, A.: Mondrian: Python Implementation for Mondrian Multidimensional K-Anonymity, GitHub, 24-Jun-2019. URL: https://github.com/qiyuangong/Mondrian (27. August 2021).

[He04] Hevner, A. R.; March, S. T.; Park, J.; Ram, S.: Design science in information systems research. In: MIS Quarterly: Management Information Systems (2004).

[Ko06] Kornblum, J.: Identifying almost identical files using context triggered piecewise hashing. In: Digital Investigation 3 (Sep. 2006) SUPPL., S. 91–97.

[Ko22] Kortum, H.; Fukas, P.; Rebstadt, J.; Eleks, M.; Nobakht Galehpardsari, M.; Thomas, O.: Proposing a Roadmap for Designing Non-Discriminatory ML Services: Preliminary Results from a Design Science Research Project. Wirtschaftsinformatik 2022 Proceedings, 2022.

[Kr57] Kreer, J. G.: A Question of Terminology. In: IRE Transactions on Information Theory 3 (1957) 3, S. 208.

[LDR06] LeFevre, K.; DeWitt, D. J.; Ramakrishnan, R.: Mondrian multidimensional K-anonymity. Proceedings - International Conference on Data Engineering, 2006, vol. 2006, S. 36.

[Le66] Levenshtein, V. I.: Binary codes capable of correcting deletions, insertions, and reversals. In: Soviet physics doklady 10 (1966) 8, S. 707–710.

[LLV07] Li, N.; Li, T.; Venkatasubramanian, S.: t-closeness: Privacy beyond k-anonymity and l-diversity. IEEE 23rd International Conference on Data Engineering, 2007, S. 106–115.

[LZ20] Lisin, N.; Zapechnikov, S.: Methods and Approaches for Privacy-Preserving Machine Learning. Advanced Technologies in Robotics and Intelligent Systems. vol. 80Springer, 2020, S. 141–148.

[Ma07] Machanavajjhala, A.; Kifer, D.; Gehrke, J.; Venkitasubramaniam, M.: L-diversity: Privacy beyond k-anonymity. In: ACM Transactions on Knowledge Discovery from Data (TKDD) 1 (Mar. 2007) 1, S. Article 3-52 Pages.

[Ma16] Malle, B.; Kieseberg, P.; Schrittwieser, S.; Holzinger, A.: Privacy Aware Machine Learning and the “Right to be Forgotten,” ERCIM News, S. 1, 26-Sep-2016.

[Mu20] Mustafa, I.; Mustafa, H.; Azar, A. T.; Aslam, S.; Mohsin, S. M.; Qureshi, M. B.; Ashraf, N.: Noise free fully homomorphic encryption scheme over non-associative algebra. In: IEEE Access 8 (2020), S. 136524–136536.

[MW04] Meyerson, A.; Williams, R.: On the complexity of optimal k-anonymity. Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, 2004, vol. 23, S. 223–228.

[Na04] National Institute of Standards and Technology, I. T. L.: Hash Functions | CSRC, NIST Information Technology Laboratory. S. 1, 2004.

[Pa16] Patki, N.; Wedge, R.; Veeramachaneni, K.; MIT Data to AI Lab: The synthetic data vault. Proceedings - 3rd IEEE International Conference on Data Science and Advanced Analytics, DSAA 2016, 2016, S. 399–410.

[Pa18] Park, N.; Mohammadi, M.; Gorde, K.; Jajodia, S.; Park, H.; Kim, Y.: Data synthesis based on generative adversarial networks. In: Proceedings of the VLDB Endowment 11 (Jun. 2018) 10, S. 1071–1083.

[Pe07] Peffers, K.; Tuunanen, T.; Rothenberger, M. A.; Chatterjee, S.: A design science research methodology for information systems research. In: Journal of Management Information Systems 24 (2007) 3, S. 45–77.

[Pe11] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E.: Scikit-learn: Machine Learning in {P}ython. In: Journal of Machine Learning Research 12 (2011), S. 2825–2830.

[Pe20] Peffers, K.; Tuunanen, T.; Gengler, C. E.; Rossi, M.; Hui, W.; Virtanen, V.; Bragge, J.: Design Science Research Process: A Model for Producing and Presenting Information Systems Research. Proceedings of the First International Conference on Design Science Research in Information Systems and Technology (DESRIST 2006), 2020, S. 83–106.

[PMJ18] Park, N.; Mohammadi, M.; Gorde, K.; Jajodia, S.; Park, H.; Kim, Y.: Data synthesis based on generative adversarial networks. In: Proceedings of the VLDB Endowment 11 (Jun. 2018) 10, S. 1071–1083.

[Re22] Rebstadt, J.; Remark, F.; Fukas, P.; Meier, P.; Thomas, O.: Towards Personalized Explanations for AI Systems: Designing a Role Model for Explainable AI in Auditing. Wirtschaftsinformatik 2022 Proceedings, 2022.

[Ro09] Roussev, V.: Building a better similarity trap with statistically improbable features. 42nd Hawaii International Conference on System Sciences, 2009, S. 1–10.

[RS04] Rowley, J.; Slack, F.: Conducting a literature review. In: Management Research News 27 (2004) 6, S. 31–39.

[RSA78] Rivest, R.; Shamir, A.; Adleman, L.: On Data Systems and Privacy Homomorphisms. In: Foundations of secure computation 4(11) (1978) 11, S. 169–180.

[RU96] Revow, M.; UCI: Adult Income Dataset. Irvine, California, 1996.

[RU96b] Revow, M.; UCI: Adult Income Dataset Description, Toronto University of Computer Science, 1996. URL: http://www.cs.toronto.edu/~delve/data/adult/adultDetail.html.

[Sc22] Schuir, J.; Anton, E.; Eleks, M.; Teuteberg, F.: Tell Me and I Forget, Involve Me and I Learn: Design and Evaluation of a Multimodal Conversational Agent for Supporting Distance Learning. Wirtschaftsinformatik 2022 Proceedings, 2022.

[Sh48] Shannon, C. E.: A mathematical theory of communication. In: The Bell system technical journal 27 (1948) 3, S. 379–423.

[SL07] Sadowsky, C.; Levin, G.: Simhash: Hash-based similarity detection, Technical Report. S. 1–10, 2007.

[SL09] Sokolova, M.; Lapalme, G.: A systematic analysis of performance measures for classification tasks. In: Information processing & management 45 (2009) 4, S. 427–437.

[Sw02a] Sweeney, L.: ACHIEVING k-ANONYMITY PRIVACY PROTECTION USING GENERALIZATION AND SUPPRESSION, International Journal of Uncertainty, Puzziness and Knowledge-Based Systems, vol. 10, no. 5, S. 571–588, 2002.

[Sw02b] Sweeney, L.: k-ANONYMITY: A MODEL FOR PROTECTING PRIVACY, International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol. 10, no. 5, S. 557–570, 2002.

[Sw97] Sweeney, L.: Guaranteeing anonymity when sharing medical data, the Datafly System. Proceedings : a conference of the American Medical Informatics Association. AMIA Fall Symposium, 1997, S. 51–55.

[Ul18] ULB Machine Learning Group: Credit Card Fraud Detection Dataset, Kaggle, 2018. URL: https://www.kaggle.com/mlg-ulb/creditcardfraud (06. August 2021).

[Vo09] vom Brocke, J.; Niehaves, B.; Simons, A.; Riemer, K.: Reconstructing the Giant : On the Importance of Rigour in Documenting the Literature Search Process. Proceedings of the European Conference on Information Systems 2009, 2009.

[VPB16] Venable, J.; Pries-Heje, J.; Baskerville, R.: FEDS: A Framework for Evaluation in Design Science Research. In: European Journal of Information Systems 25 (2016) 1, S. 77–89.

[WE18] Wagner, I.; Eckhoff, D.: Technical privacy metrics: A systematic survey. In: ACM Computing Surveys 51 (2018) 3, S. 1–45.

[WHZ13] Wu, L.; He, H.; Zaïane, O. R.: Utility enhancement for privacy preserving health data publishing. ADMA 2013: Advanced Data Mining and Applications, 2013, vol. 8347 LNAI, no. PART 2, S. 311–322.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: Management Information Systems Quarterly 26 (2002) 2, S. xiii–xxiii.

[Xu19] Xu, L.; Skoularidou, M.; Cuesta-Infante, A.; Veeramachaneni, K.: Modeling Tabular data using Conditional GAN. 33rd Conference on Neural Information Processing Systems, 2019, vol. 32, S. 1–15.

[YPB14] Yi, X.; Paulet, R.; Bertino, E.: Homomorphic Encryption. Homomorphic Encryption and Applications. 1st ed., vol. 1Springer, Cham, Melbourne, Australia, 2014, S. 27–46.

Issues of Verifying Anonymity: An Overview

Sarah Stummer[1]

Abstract: The processing of anonymous data has great potential for research, economy, and society. However, due to legal ambiguities regarding anonymization and the verification of anonymity, its potential is not exploited yet. This paper addresses the chances and issues regarding the concept of anonymization and discusses what needs to be clarified to enable organizations to verify whether data are anonymized in a GDPR-compliant way. Based on this discussion, the paper proposes the implementation of a metrics system as a possible preliminary solution to solve the issues with anonymization and the verification of anonymity.

Keywords: Anonymity, Anonymization, De-anonymization, GDPR, Personal data.

1 Introduction

Over the last years, technology advanced rapidly, and the world became more connected as well as data-driven. Connecting to friends, family members or colleagues via social media, instant messaging services or web conferencing tools, cooking and eating dinner with smart devices assisting or training with a smart watch recording the heart rate, training-progresses etc. – all these and numerous more activities nowadays include technology which allows to connect with other people, systems or applications. As a result of the technological advancement and increasing connectivity, a large amount of data are processed and data are becoming more and more significant.

However, even though the data are required to offer and improve the mentioned services, and therefore, required to serve the interests of individuals, they often refer to a certain natural person and thus are personal data. Hence, the technological advancement and increasing connectivity can also pose privacy risks for individuals. In particular, individuals, their behavior and interests as well as their connection to other individuals become traceable. This can lead, e.g., to identity theft, financial or social disadvantages [Da18], social engineering as well as decreasing personal privacy.

As there is such a tension between the significance of data and the added value data create on one hand and privacy risks that are posed by the use of personal data on the other hand,

[1] Fraunhofer Institute for Secure Information Technology SIT | ATHENE – National Research Center for Applied Cybersecurity, Cloud Computing, Identity & Privacy, Rheinstr. 75, 64295 Darmstadt, Germany, sarah.stummer@sit.fraunhofer.de, https://orcid.org/0000-0003-4015-4429.

doi:10.18420/inf2022_17

the European legislator tightened legal obligations in connection to the processing of personal data as well as monetary consequences in the event of an infringement of data protection obligations. As a result, companies must comply with multiple data protection obligations in order to process personal data and make use of the added value provided by data processing. In addition, limits are set insofar as the processing of personal data under certain conditions is not permitted.

To create a balance between the added value created by data processing and the protection of individual privacy as well as to evade the scope of data protection obligations, where natural persons do not need to be identified or identifiable, data anonymization and the processing of information that not (or no longer) relate to an identified or identifiable natural person ('anonymous data') may be a measure. However, due to ambiguities and issues with the concept of anonymization, its potential is not exploited, and the added value of anonymization is not used in practice yet. This paper discusses the chances and issues regarding the concept of anonymization and points out what needs to be clarified to anonymize data and verify anonymity in a GDPR-compliant way. Based on this discussion, the paper proposes the implementation of a metrics system as a possible solution to enable organizations to assess anonymity and thus, to create a positive sum between privacy and the significance of data.

2 Related Work

Anonymization is the subject of multiple publications in the technical and legal literature. This reaches from publications on the relevance of anonymization to publications on the legal and technical analysis of the terms 'anonymous' and 'anonymization' to issues regarding anonymization and anonymization techniques.

Relevance of anonymization Since anonymization is considered a measure to reduce the data subjects' risk for harms [An19], it is relevant for general data protection. Besides that, the potential of anonymization is seen e.g. for research projects and business models not requiring references to natural persons [Bu20], [Gi21], especially for social science surveys, statistical analyses, and the development of new services or products [Ar14]. A concrete application scenario, where anonymization can have an added value (inter alia by enabling the identification of needs, e.g., for medical care or parking spaces in residential areas) is the smart city [ST21a].

Concept of anonymization From a legal point of view the terms 'anonymous' and 'anonymization' are analyzed and discussed in multiple publications, especially with regards to whether the term 'anonymous' is to be understood absolute or relative (e.g. in [Bi20], [Gi21], [Ma16]). Additionally, multiple supervisory authorities have commented on anonymization and its legal requirements ([Eu20], [Ar14], [An19]). From an

interdisciplinary perspective, [Hö19] addresses differences and similarities in the legal and technical concept of anonymization as well as their implications for practice. Accordingly, the legal and technical concept of anonymization especially differ in the group of persons for whom data must be anonymous: Whereas from a legal perspective data must be anonymous for everyone, including the controller, from a technical perspective data must only be anonymous for external attackers. [AE20] propose a procedure for assessing identifiability, respectively, anonymity by the consideration of the five risk dimensions according to the Five Safes framework (safe projects, safe people, safe data, safe settings and safe outputs [DRW16]).

Issues with anonymization Especially the legal discussions regarding anonymization often lead to the conclusion that the terms and their conditions are not specific enough for anonymization being used in practice [WBH19], [MBH18], [ST21b]. Reasons for this are seen, inter alia, in the expansion of the term 'personal data' on data that initially do not appear to be personal [Oh10], missing processes for anonymization and a lack of ability to assess whether data are anonymous [ST21b]. Moreover, there are multiple publications regarding insufficient anonymization and de-anonymization (e.g. [NS08], [MHVB13], [Bo13]). For instance, [MHVB13] watched geolocation data of 1,5 million people and proved the possibility of identifying 95% of the people observed with only four location-time-points. As a result, for organizations it is very hard to determine when data are actually anonymized.

3 Chances of Anonymization

When personal data are processed, privacy regulations, such as the GDPR, become applicable. As a result, controllers must comply with data protection obligations. Under the GDPR, these obligations especially include the principles relating to the processing of personal data stated in Art. 5 (1) GDPR. Accordingly, the processing of personal data must be transparent and fair and requires a legal basis in order to be lawful. Additionally, the processing of personal data only is allowed for specified, explicit, and legitimate purposes and should be adequate, relevant, and limited to what is necessary for achieving the purposes of the processing. Further processing only is allowed in a manner that is compatible with those purposes the personal data originally were collected for.

Following from the principles of data protection in terms of Art. 5 (1) GDPR, it is not always lawful to process personal data. This may be the case when the purpose of the desired processing cannot be considered fair, for instance when personal data collected by a smart voice assistant (e.g. search requests, shopping lists, and behavior of a certain user) are to be disclosed to other parties (e.g. manufacturers) to show the demand for specific products (e.g. medicine). This may also be the case when the controller desires to process

data for another purpose than what they were collected for, for instance for research purposes, for the training of an artificial intelligence or for the improvement of the collectors' products and services, and/or because reference to a natural person is not required to achieve the desired purposes. Also, the processing of personal data could be unlawful due to the lack of a legal basis for processing.

To counter these legal barriers and enable organizations to achieve their desired purposes, when reference to a natural person is not required, anonymization and the processing of anonymous data might be a suitable solution.

Anonymization is the process of rendering personal data anonymous, so that the reference to a natural person of the data is removed [PP21]. Therefore, it enables organizations to make use of data without having to comply to data protection obligations and without posing a risk for privacy. Additionally, it potentially enables the processing of data for purposes for which it could not be processed if it was personal data. In this context, anonymization might also be an alternative to erasure (affirmative: [ESS19], [St20]; negative: [Ro21]).

The potential of anonymization is also seen in practice. This in particular is shown by an empirical interview of 30 smart city participants, from which 28 interviewees stated that anonymous data can create added value for smart cities [ST21a]. Such added value can be reached, for instance, by

- data exchange between public agencies and social networking platforms (e.g. to improve the responds to disaster alerts),
- data sharing between e.g. educational institutions, public health departments, and health care providers (e.g. to assist in the mitigation of spreading contagious diseases) and
- data exchange between health care providers, research institutions, and civil service offices (e.g. "to identify areas in communities where certain diseases are occurring more frequently, so that the cause can be researched and help can be offered") [ST21a].

Moreover, anonymization becomes increasingly important in research and economy [Bu20], [Gi21], e.g. for training, testing and evaluating artificial intelligence [RG21].

4 Issues regarding the Concept of Anonymization

Even though anonymization has great potential and can be an appropriate solution to create a balance between privacy and data use, there are certain issues with the concept of anonymization – especially regarding the verification of anonymity – that must be concretized for organizations to make use of its potential. These issues will be discussed in the following.

4.1 Boundaries between Personal Data and Anonymous Data

Firstly, there is an obscurity in the boundaries between personal and anonymous data as well as the allocation of data as personal or anonymous.

Legal definitions of the GDPR

According to Art. 4 (1) GDPR personal data are any information relating to an identified or identifiable natural person (so called 'data subject'), whereas an identifiable natural person is one who can be identified, directly or indirectly. Hence, for data to be considered personal data it is not required that the data subject is identified directly through direct identifiers, such as passport numbers, social security numbers, unique customer numbers or – depending on the context – names and addresses [EM15] (identifying personal data). Rather, the possibility of identifying a natural person through so-called quasi-identifiers is sufficient as well (identifiable personal data). Quasi-identifiers are datasets consisting of attributes which individually do not identify a natural person but in their combination become identifying. For instance, information on the date of birth, gender, and zip codes in their combination can be considered a quasi-identifier, depending on the context [Sw00].

A special form of personal data are pseudonymized data. The term pseudonymized data is not directly defined within the GDPR, however, from the definition of pseudonymization according to Art. 4 (5) GDPR it can be concluded that pseudonymized data according to the GDPR are personal data that can no longer be attributed to a specific data subject without the use of additional information, which is kept separately and is subject of technical and organizational measures.

Regarding anonymous data the European legislator did not introduce a definition. Instead, the term only is mentioned in the Recitals of the GDPR, especially in Recital 26. Accordingly, the principles of data protection only apply to information concerning an identified or identifiable natural person and thus, are not applicable to anonymous data. Hence, anonymous data are differentiated from personal data insofar as anonymous information is information not relating to an identified or identifiable natural person or

personal data rendered anonymous so that the data subject no longer is identifiable. Anonymous data therefore can be understood as the opposite of personal data [Ro21].

Differences between personal data and anonymous data

Based on the legal definitions of the terms ‘anonymous data’ and ‘personal data’, it becomes clear that the terms differ in their ability to identify a natural person as well as in their degree of identifiability. In this regard, the boundaries between the terms seem to be clear: Personal data generally allow to identify a natural person and therefore, exhibit a higher degree of identifiability than anonymous data, where the identification of a specific natural person is not possible (anymore). However, as shown in Figure 1, with regards to personal data, it needs to be further distinguished between identifying personal data and identifiable personal data. Since identifying personal data allow the identification of a natural person directly, whereas identifiable personal data only allow identification in combination with other information and anonymous data do not allow identification at all, the degree of identifiability increases with each layer shown in Figure 1.

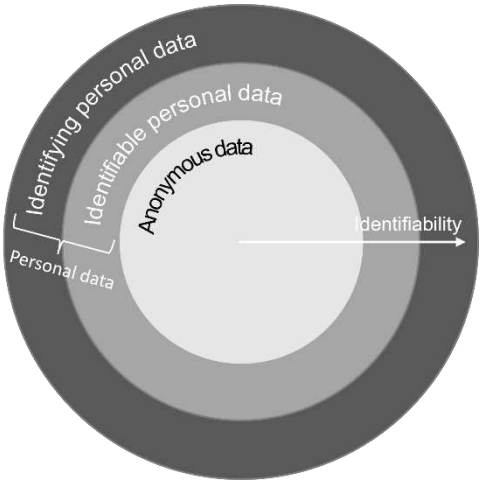


Figure 1: Abstract view on the boundaries between personal data and anonymous data as well as their sub-categories based on their ability to identify a specific natural person.

Assessing whether data are personal data or anonymous data

Though, the legal definitions and the differences in the terms are clear, the boundaries between the terms are blurring, when it comes to assessing whether data refer to an identified or identifiable natural person and determining if a dataset is anonymous or not.

First, whether personal data are identifying or identifiable depends on the specific context, especially the size of the group of eligible natural persons (‘anonymity set’ [PKS05]) [Ar07]. For instance, in a group of 50 natural persons let the persons‘ eye colors be given and distributed like this: 10x green, 15x blue, and 25x brown. In this case, the attribute ‘blue eyes’ alone would not identify one specific natural person but would be applicable to fifteen persons. Hence, for identifying a specific natural person, the combination of the eye color with other attributes – such as date of birth, hair color or initials – would be required. However, if the group would only consist of 7 individuals, with only one person having blue eyes (e.g. 2x green eyes, 4x brown eyes, 1x blue eyes), the attribute ‘blue eyes’ would directly identify a natural person and thus, would qualify as identifying personal data. Therefore, when assessing whether data are directly or indirectly identifying, the context of the processing, especially with regards to the anonymity set, must be taken into account [Ar07]. A categorization of data as identifying personal data or identifiable personal data based solely on the data is not sufficient.

It becomes even more complex when it further is assessed whether data are personal data at all or if they qualify as anonymous data. Especially the boundary between anonymous data and identifiable personal data is blurring. This is because pursuant to Recital 26 GDPR different aspects need to be considered when assessing whether a natural person is identifiable and thus, data are personal or anonymous. Accordingly, when determining whether a natural person is identifiable, all the means reasonably likely to be used by the controller or by another person to identify the natural person should be taken into account. The likelihood of means used is depending on all objective factors, such as the costs, the amount of time as well as available identification-technologies. Hence, instead of a clear distinction between identifiable personal data and anonymous data, the European legislator provides different factors that must be considered when determining whether data have a reference to a natural person or are anonymous. As a result, from a legal point of view, it is unclear when data are anonymous. Instead, there is a fine line between (identifiable) personal data and anonymous data, without a specification on where exactly this line runs (see Figure 2).

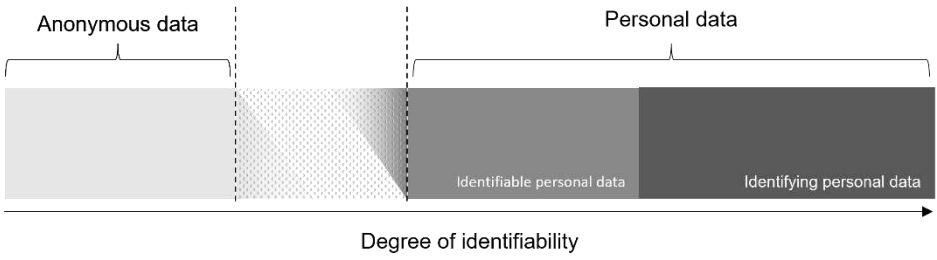


Figure 2: Blurring boundaries between anonymous data and personal data.

In order to create legal certainty for the organizations anonymizing data and to enable them to make use of anonymization and its potential to create added value, it therefore must be concretized where the line between personal data and anonymous data runs, which requirements must be met to consider data sufficiently anonymous and thus, under which conditions data are not personal data but qualify as anonymous.

Issues to solve:

Where does the boundary between anonymous data and personal data lie? When are data anonymous?

Which requirements must be met to consider data sufficiently anonymous?

4.2 Possibility of Anonymity

Even though it is not conclusively resolved where the boundary between anonymous data and personal data lies and when data are anonymous from a legal point of view, anonymity and techniques to render personal data anonymous (anonymization techniques) are discussed sufficiently.

Anonymization techniques

From a technical perspective, there are numerous techniques to anonymize data. The foundation for anonymization generally is the removal or masking of direct identifiers to ensure that direct identification is not possible anymore. However, to fully anonymize data, it furthermore requires that indirect identification of a natural person is ruled out practically. Therefore, additional techniques to generalize and/or randomize data must be applied to such an extent that natural persons are not identifiable anymore. Through generalization, data are abstracted and generalized in such a way that singling out a natural person becomes less likely. Through randomization, data are distorted in such a manner that their connection to a natural person is removed [Ar14].

Insufficient anonymization and de-anonymization

However, as multiple studies have shown, if anonymization techniques are applied naively or insufficiently, identifying natural persons may still be possible and thus, the resulting data may not be anonymous. For instance, [NS08] analyzed attributes of Netflix' customer database with more than 100 million movie ratings from a scale of 1 to 5. Except for the rating and its point of time, every identifier was removed. Additionally, randomization was conducted. However, the researchers still were able to identify 99% of the records within the dataset with only eight ratings and the dates of their delivery (with a 14-day error). 68% of the users even could be identified with only two ratings and the time of

their delivery, known to within three days. Another example for the identifiability of data is given by [MHVB13], who watched geolocation data of 1,5 million people for a period of 15 months and found out that it requires only four location-time-points to identify 95% of the people observed.

The possibility of identifying a natural person is amplified by the technological advancement and interconnectivity over the internet – both resulting in an increasing amount and accessibility of data. For instance, when contacting friends, family members or other contacts, in multiple cases, direct messenger services, such as WhatsApp or Instagram Direct Messaging, are used. Additionally, personal meetings and leisure activities often are recorded by phone cameras and shared – potentially connected to a geolocation – with other individuals via social media. Another example are household activities: An increasing number of households use smart home devices, such as vacuum robots, smart refrigerators, smart washing machines and voice assistants (e.g. Alexa), which support by turning the light on or setting timers but also by looking for recipes, answering questions or ordering goods. All these and numerous more activities process data and therefore result in an increasing amount and accessibility of data allowing inferences to individuals. The question therefore is whether in an era of big data, where by the use of services a large amount of data is collected, information is shared in the internet and a digital trace is left by individuals, anonymity is even possible.

Hence, it needs to be assessed whether the legal requirements for anonymity, especially due to the increasing accessibility of data, can be met and thus, whether sufficient anonymity is possible from a legal point of view. In addition to the question of whether anonymity generally is possible, it needs to be discussed whether and to what extent the (loss of) value of data can be considered when it comes to assessing anonymity. Usually, the more the data are generalized and randomized, the less value they have. However, since anonymization is not only carried out to increase privacy, but also to use the anonymized data for certain purposes, the resulting data must be meaningful and sufficient to fulfill the desired purposes. Otherwise, the anonymized data loses their value, and their processing becomes useless. Thus, it is questionable whether anonymity is possible without losing the data's information value and to what extent such considerations can be taken into account when it comes to anonymization and verifying anonymity (e.g. by requiring anonymization only to such an extent to which the resulting data still are useful).

Issues to solve:

Is anonymity possible from a legal point of view?

Can loss of information value be considered when assessing anonymity, so that data must only be anonymized to an extent to which they still are useful?

4.3 Relevance of the Context and Risk

Connected to the possibility of anonymity and the blurring boundaries between anonymous and personal data, it is also questionable whether and to what extent the context of the processing as well as the risk that is connected to the processing can be considered when assessing anonymity.

For instance, when an anonymized dataset is only to be used internally within the anonymizing organization and appropriate measures to ensure confidentiality are to be taken, it is also only possible for the anonymizing organization to try carrying out de-anonymization. In this case, perhaps less means to identification need to be considered when determining whether data are anonymous or not. The same applies for cases, in which data is collected in a form that is anonymous for the data controller, even though it cannot be ruled out that other parties could be able to identify a natural person.

An argument for considering risk and context is the jurisdiction of the European Court of Justice. Accordingly, all legal means enabling a party to identify a natural person [Eu16] (including the involvement of third parties which have the information on identity and are legally compelled to provide this information [Ge17]) must be considered when assessing whether data are personal or anonymous. From this follows that only the lawful means of the parties having access to the anonymized dataset must be taken into account. Another argument for considering risk and context are the requirements of the GDPR in connection to technical and organizational measures, especially the GDPRs' risk-based approach. Art. 24, 25, 32 GDPR require the controller to take technical and organizational measures in order to implement data protection principles stated in Art. 5 (1) GDPR as well as to ensure an appropriate level of security. When taking such measures, the controller shall, inter alia, take the risk as well as the context and purposes of processing into account. Since data anonymization removes the data's reference to a natural person and therefore ensures data minimization in terms of Art. 5 (1) (c) GDPR, it serves as a measure to implement data protection principles [Go18].

However, in contrast to this is that data anonymization results in data protection obligations not being applicable anymore. Thus, it is questionable whether such a risk- and context-based approach can be followed (affirmative e.g. [Ag21], [An19]; negative e.g. [PP21]). This consideration also is connected to the dispute of whether anonymity must be absolute or relative. According to the absolute theory of identifiability, data only can be considered anonymous if it is impossible for anyone to identify a natural person. The relative theory of identifiability, on the other hand, considers data anonymous when identification due to all objective factors, such as the costs and the amount of time required for identification, reasonably is not likely (even though theoretically it was possible) [SHS19]. Regarding this dispute, the GDPR is not unambiguous. In contrast, Recital 26 GDPR contains elements of the absolute theory as well as elements of the relative theory:

In favor of the relative theory of identifiability is that for the determination of whether a natural person is identifiable, only the means reasonably likely to be used for the identification of natural person should be taken into account. On the other hand, the GDPR addresses all the means used by the controller or another person which is characteristic for the absolute theory of identifiability. Thus, the European legislator seems to generally support the relative theory but restricts it through elements of the absolute theory [SHS19], [Ma16].

Issues to solve:

Can a risk- and context-based approach be followed when assessing anonymity?

(To what extent) must anonymity be absolute or relative?

4.4 Consequences of De-Anonymization

As already mentioned, anonymity is not static but can change dynamically [Ma16], [HW19]. Hence, in the past there have been multiple examples where alleged anonymity has been reversed (so called 'de-anonymization'), so that conclusions to a specific natural person have been possible (again) and natural persons became (re-)identifiable.

De-anonymization can be caused by anonymization techniques being applied naively or insufficiently, by new, faster or cheaper technologies [HW19], [HW20] as well as by increased accessibility of data [Bo13].

As a result of de-anonymization, data becomes personal data (again) and data protection obligations must be fulfilled when the data are processed. However, since de-anonymization can also result in natural persons being identifiable rather than identified (for instance in cases where in an insufficient anonymized dataset only a few natural persons are identified and it cannot be ruled out that more natural persons are identifiable), data protection obligations which require the identification of a natural person cannot always be fulfilled. The questions therefore are, what the consequences of insufficient anonymization or de-anonymization are and whether and to what extent legal obligations of the GDPR must be fulfilled when natural persons become re-identifiable.

Issue to solve:

Must data protection obligations be fulfilled in consequence of data-de-anonymization and to what extent?

5 Approach for a solution

As we have seen, there are multiple open issues regarding the concept of anonymization, all of which relate to clarifying whether a particular dataset is anonymous or not, and how to deal with a dataset that was assumed to be anonymous, but which has been found as being personal data by a detailed analysis of the dataset. As a result of these issues, it can hardly be determined whether a dataset is anonymous. The chances of anonymization therefore are not exploited in practice yet, even though many important added values for society and economy could be achieved.

To enable organizations to assess and verify anonymity of a dataset and hence, enable them to exploit the chances of anonymization, one possible way could be the development and use of a metrics system. Metrics are mathematical functions that allow to systematically measure the adherence to requirements and to compare the as-is state with the target state [AS12]. Metrics therefore could also be used to assess and verify whether the requirements for anonymity are fulfilled by an anonymized dataset.

The development and use of a metrics system for the assessment and verification of anonymity, however, first requires the mentioned issues to be solved. Therefore, anonymity and its requirements must be analyzed from an interdisciplinary view considering law and technology. In particular, it must be assessed under which conditions means for identification are reasonably likely to be used, to what extent a risk- and context-based approach can be followed when the likelihood of means to be used for identification is assessed and where the boundary between personal data and anonymous data lies. Based on the findings of the analysis, the legal requirements for anonymity become more concrete and verifiable requirements and measures for anonymization can be derived. These verifiable measures then serve as the foundation for a set of metrics, through which anonymity can be assessed and verified.

Since anonymity can change dynamically, organizations furthermore should be provided with a guideline addressing the privacy-compliant handling of non-anonymous data. Therefore, the legal consequences of de-anonymization need to be analyzed (possibly depending on the use case) and a procedure for organizations to handle non-anonymous data needs to be elaborated.

This contribution therefore proposes the following approach:

1. Answering the mentioned issues regarding the concept of anonymization with an interdisciplinary perspective considering law and technology.
2. Deriving verifiable metrics from the results of (1).

3. Implementing a metrics system to verify the degree of anonymity of a dataset as well as a guideline that shows the privacy-compliant handling of non-anonymous data, based on the results of (2).

6 Conclusion and Outlook

As there are multiple issues with the determination of anonymity, which are not sufficiently clarified yet, even though anonymization has great potential for society, research, and economy and is increasingly important for the privacy of individuals, its potential is not exploited yet.

To enable organizations to benefit from the added value anonymization and the processing of anonymous data can create, the issues mentioned in this paper need to be clarified and a process for verifying anonymity must be elaborated. Initial indications for solving some of the mentioned issues could be provided by the position guideline for anonymization and pseudonymization, which the European Data Protection Board announced for 2021/2022 [Eu21]. Additionally, I will work on answering the mentioned issues with an interdisciplinary perspective considering law and technology to elaborate a metrics system for the assessment and verification of anonymity.

This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[AE20] Arbuckle, Luk; El Emam, Khaled: Building an Anonymization Pipeline. Creating Safe Data, O'Reilly Media, 2020.

[Ag21] Agencia Española de Protección de Datos: 10 Misunderstandings related to Anonymisation. At: https://edps.europa.eu/system/files/2021-04/21-04-27_aepd-edps_anonymisation_en_5.pdf, 2021.

[An19] An Coimisiún um Chosaint Sonraí: Guidance Note: Guidance on Anonymisation and Pseudonymisation. At: https://www.dataprotection.ie/sites/default/files/uploads/2020-09/190614%20Anonymisation%20and%20Pseudonymisation.pdf, 2019.

[Ar07] Article 29 Data Protection Working Party: Opinion 4/2007 on the concept of personal data (WP 136). At: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf, 2007.

[Ar14] Article 29: Data Protection Working Party: Opinion 05/2014 on Anonymisation Techniques (WP 216), Brüssel, 2014.

[AS12] Ammann, Franz-Ernst; Sowa, Aleksandra: Systematische Entwicklung von Metriken zur Beurteilung der Datensicherheit. In: Datenschutz und Datensicherheit (DuD), p. 247-251, 2012.

[Bi20] Bischoff, Claudia: Pseudonymisierung und Anonymisierung im Rahmen klinischer Prüfungen von Arzneimitteln (Teil I). In: Pharma Recht (PharmR), p. 309-315, 2020.

[Bo13] Bohannon, John: Genealogy Databases Enable Naming of Anonymous DNA Donors, Science, 339 (6117), p. 262, 2013.

[Bu20] Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit: Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche. Bonn, 2021.

[Da18] Datenschutzkonferenz: Kurpapier Nr. 18 Risiko für die Rechte und Freiheiten natürlicher Personen. At: https://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_18.pdf, 2018.

[DRW16] Desai, Tanvi; Ritchie, Felix; Welpton, Richard: Five Safes: designing data access for research. Economics Working Paper Series. University of the West of England, Bristol, England. Faculty of Business and Law, 2016.

[EM15] El Emam, Khaled; Malin, Bradley: Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington D.C.: National Academies Press, 2015.

[ESS19] Enzmann, Matthias; Selzer, Annika; Spychalski, Dominik: Data Erasure under the GDPR – Steps towards Compliance. In: European Data Protection Law Review (EDPL), p. 416-420, 2019.

[Eu16] European Court of Justice, C-582/14, 19.10.2016.

[Eu20] European Data Protection Board: Guidelines 04/2020 on the use of location data and contact tracing tools in the context of the COVID-19 outbreak. At: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_20200420_contact_tracing_covid_with_annex_en.pdf, 2020.

[Eu21] European Data Protection Board: EDPB Work Programme 2021/2022. At: https://edpb.europa.eu/system/files/2021-03/edpb_workprogramme_2021-2022_en.pdf, 2021.

[Ge17] German Federal Court of Justice, VI ZR 135/13, 16.05.2017.

[Gi21] Gierschmann, Sybille: Gestaltungsmöglichkeiten durch systematisches und risikobasiertes Vorgehen – Was ist schon anonym? Planung und Bewertung der Risiken der Anonymisierung. In: Zeitschrift für Datenschutz (ZD), p. 482-486, 2021.

[Go18] Gola, Peter: Datenschutz-Grundverordnung. 2. Auflage, C.H. Beck, München, 2018.

[Hö19] Hölzel, Julian: Differential Privacy and the GDPR. In: European Data Protection Law Review (EDPL), p. 184-196, 2019.

[HW19] Hornung, Gerrit; Wagner, Bernd: Der schleichende Personenbezug: Die Zwickmühle der Re-Identifizierbarkeit in Zeiten von Big Data und Ubiquitous Computing. In: Computer und Recht (CR), p. 565-574, 2019.

[HW20] Hornung, Gerrit; Wagner, Bernd: Anonymisierung als datenschutzrelevante Verarbeitung? Rechtliche Anforderungen und Grenzen für die Anonymisierung personenbezogener Daten. In: Zeitschrift für Datenschutz (ZD), p. 223-228, 2020.

[Ma16] Marnau, Ninja: Anonymisierung, Pseudonymisierung und Transparenz für Big Data: Technische Herausforderungen und Regelungen in der Datenschutz-Grundverordnung. In: Datenschutz und Datensicherheit (DuD), p. 428–433, 2016.

[MBH18] Marnau, Ninja; Berrang, Pascal; Humbert, Mathias: Anonymisierungsverfahren für genetische Daten, In: Datenschutz und Datensicherheit (DuD), p. 83-88, 2018.

[MHVB13] de Montjoye, Yves-Alexandre; Hidalgo, César A.; Verleysen, Michel; Blondel, Vincent D.: Unique in the Crowd: The privacy bounds of human mobility. In: Scientific Reports 3:1376, p. 1-5, 2013.

[NS08] Narayanan, Arvind; Shmatikov, Vitaly: Robust de-anonymization of large sparse datasets. In IEEE Symposium on Security and Privacy, p. 111-125, 2008.

[Oh10] Ohm, Paul: Broken promises of privacy: Responding to the surprising Failure of Anonymization. In: UCLA Law Review (57), p. 1701–1777, 2010.

[PKS05] Pfitzmann, Andreas; Köhntopp, Marit; Shostack Adam et. al: Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management – A Consolidated Proposal for Terminology. At: https://dud.inf.tu-dresden.de/literatur/Anon_Terminology_v0.22.pdf, 2005.

[PP21] Paal, Boris P.; Pauly, Daniel A.: Datenschutz-Grundverordnung Bundesdatenschutzgesetz. 3. Auflage, C.H. Beck, München, 2021.

[RG21] Roßnagel, Alexander; Geminn, Christian L.: Vertrauen in Anonymisierung. Regulierung der Anonymisierung zur Förderung Künstlicher Intelligenz. In: Zeitschrift für Datenschutz (ZD), p. 487-490, 2021.

[Ro21] Roßnagel, Alexander: Datenlöschung und Anonymisierung – Verhältnis der beiden Datenschutzinstrumente nach DS-GVO. In: Zeitschrift für Datenschutz (ZD), p. 188-192, 2021.

[SHS19] Simitis, Spiros; Hornung, Gerrit; Spiecker gen. Döhmann, Indra: Datenschutzrecht. 1. Auflage, C.H. Beck, 2019.

[St20] Stürmer, Verena: Löschen durch Anonymisieren? In: Zeitschrift für Datenschutz (ZD), p. 626-631, 2020.

[ST21a] Selzer, Annika; Timm, Ingo: Chances and Limitations of Personal and Anonymized Data Processing. In: INFORMATIK 2021. Gesellschaft für Informatik, Bonn, p. 773-787, 2021.

[ST21b] Selzer, Annika; Timm, Ingo: Potenziale anonymer Datenverarbeitungen nutzen. In: Datenschutz und Datensicherheit (DuD), p. 816-820, 2021.

[Sw00] Sweeney, Latanya: Simple Demographics Often Identify People Uniquely. In: Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh, 2000.

[WBH19] Winter, Christian; Battis, Verena; Halvani, Oren: Herausforderungen für die Anonymisierung von Daten. Technische Defizite, konzeptuelle Lücken und rechtliche Fragen bei der Anonymisierung von Daten. In: Zeitschrift für Datenschutz (ZD), p. 489-493, 2019.

The status-quo of companies' data privacy and security communication: An ethical evaluation and future paths

K. Valerie Carl [1]

Abstract: Advancing digital technologies and the omnipresent digitalization bear chances but also threats for companies and consumers. Especially data privacy and security risks remain one key concern of consumers. However, fulfilling legally binding requirements is not sufficient anymore for many consumers. Instead, consumers expect companies to behave ethically responsible and voluntarily assume more responsibilities in the digital context, particularly related to data privacy and security. Broader approaches, like the emerging concept of Corporate Digital Responsibility, make it possible to see responsibility with regard to data privacy and security in the broader context of a company's digital responsibilities and thus to develop a more holistic understanding. However, responsible behavior alone is not enough; rather an adequate communication is the evaluation basis of consumers. Accordingly, this work-in-progress evaluates current corporate communication regarding data privacy and security from an ethical view, thus illustrating best practices and paving future paths for corporate communication.

Keywords: data privacy, data security, company communication, best practices, communication guidelines, Corporate Digital Responsibility.

1 Introduction

The omnipresent digitalization and associated advanced digital technologies, products, and services transform private and professional lives. These advancements allow networks of devices to communicate via the Internet, exchange data, and to automate processes and tasks without human interferences. In this context, advancements in Artificial Intelligence (AI) and interconnectedness present an essential element of a plethora of products, services, and systems. Accordingly, a company's handling of the collected data is gaining in importance for, e.g., consumers and society; particularly since risks related to these activities are one of the key concerns of consumers [e.g., Ma86, MZH17]. Especially recently, and reinforced once again by the COVID-19 pandemic, consumers' awareness of data privacy and security has increased and, at the same time, so have their demands on companies to behave ethically and voluntarily exceed the minimum legally required responsibilities [IP18]. Nowadays, it is no longer sufficient for companies to comply with legal standards [Mi22]. Rather, the requirements for ethical behavior and corporate communication have increased to make ethical behavior visible. Hazards related to

[1] Goethe University Frankfurt/Main, Chair of Information Systems and Information Management, Theodor-W.-Adorno-Platz 4, D-60323 Frankfurt am Main, Germany, kcarl@wiwi.uni-frankfurt.de, https://orcid.org/0000-0003-4655-1046

doi:10.18420/inf2022_18

unethical bahvior result in economic, ethical, and legal issues for companies and consumers alike [Ba19]. Hence, data privacy and security activities of companies have the ability to positively or negatively influence a consumer's perception of a company, thus also technology adoption, with a lasting effect [Lu02]. Hence, it is of tremendous importance to direct companies' activities towards consumer requirements in this field. However, consumers are not aware of many activities. Rather they evaluate corporate communication and (mostly negative) incidents related to data privacy and security. Therefore, it is crucial for companies to align internal activities with their communication to exploit the full potential of additional efforts.

Still, firms struggle to keep up with the requirements for more ethical communication and behavior due to the lack of guidelines and best practices. Yet, standards guide socially responsible behavior (i.e., Corporate Social Responsibility (CSR)), however, with regard to data privacy and security, without concrete guidance for companies and current standards lack an adequate emphasis on this distinct topic [CZH22]. Nevertheless, there is a vast amount of research addressing data privacy and security in general [e.g., BC11, HH18] and its effect on technology adoption [e.g., Lu02]. Still, research lacks the link to an ethical evaluation of corporate activities and their communication in the digital era. One evolving concept aiming at the support of ethical digital behavior of companies is the concept of Corporate Digital Responsibility (CDR). This concept allows for a broader perspective on responsibilities related to (digital) corporate activities (e.g., data privacy and security activities). CDR and CSR both sum up to the broader concept of Corporate Responsibilities and address different aspects of a broader responsibility understanding. CSR targets social and economic impacts of corporate activities [MR02], whereas CDR puts digital responsibilities in the focus. Digital responsibilities relate to the exploitation of opportunities and the handling of potential negative outcomes of digitalization [Lob21, Mi21]. The concept of CDR subsumes yet rather isolated aspects of digital responsibilities in the bigger picture to ensure a more holistic understanding of corporate responsibilities in the digital era, aiming at enhanced consumer trust in corporate activities. Thus, the concept provides a framework not only for ethically responsible corporate activities but also for making them visible to consumers through appropriate corporate communication. In this way, communication regarding data privacy and security activities can benefit from putting associated responsibilities into the larger context to support consistent communication and enhanced consumer trust.

Nevertheless, CDR and its practical implementation are still in its infancy [Lob21] although consumers already expect ethically responsible behavior from firms. By now, we can observe an increasing discourse in research [e.g., Lob21], practice [e.g., He21], and from regulatory entities [e.g., Th17] related to the concept of CDR. However, the focus lies currently on developing a common understanding of CDR and its scope [e.g., He21, Lob21, Mi21]. Yet, concrete guidance for ethically responsible company communication, especially directed at data privacy and security activities, lacks. Therefore, primarily small and medium sized enterprises (SME) struggle when fostering a more ethical communication due to missing guidelines and best practices. Hence, it is of tremendous importance to provide further guidance especially for SME. In general, individual

consulting by consulting firms is not feasible for ethical communication in most SME unlike for larger companies. Thus, this study aims to close this gap and provides guidance on the ethically responsible communication related to data privacy and security by pointing out possible best practices and positive examples.

To achieve this goal, this study evaluates companies' data privacy and security communication to propose some best practices and to indicate future paths for a more ethical communication. Section two introduces data privacy and security responsibilities. Subsequently, the next sections present the study design and the results of the study, before drawing a conclusion and presenting an outlook and this study's implications. To derive best practices and therefore guidance on an ethical communication, the focus on business-to-consumer (B2C) companies compared to business-to-business (B2B) companies is necessary. B2C and B2B companies differ in terms of their stakeholders' demands regarding behavior and corporate communication, which makes them harder to compare. Accordingly, this study initially focuses on B2C companies for the development of best practices. B2C companies face communication needs of a variety of stakeholders (e.g., employees, consumers, society). Since consumers are increasingly sensitive to ethical behavior or the lack thereof, we selected consumers as the addressees of corporate communication. Hence, this study is concerned with B2C companies and specifically with their communication directed at consumers.

2 Data Privacy and Security Responsibilities

In general, data privacy addresses consumers' control over data usage, including but not limited to storing, processing, and forwarding of consumer data. By contrast, data security describes the protection of consumer data against (virtual) attacks and other threats [BC11]. Hence, in the digital economy, the monetization of consumer data, data use, possible control over it, and information security are in the focus of consumer concerns [e.g., BC11, We67]. Regulations define the minimum requirements for data privacy and security activities in many countries (e.g., for companies to which the GDPR applies). Compliance with these regulations does not stand out positively whereas non-compliance can have severe financial and legal consequences [e.g., GS09]. Against this background, companies have to exceed legal regulations to positively influence a consumer's perception of a company [Ha07]. Voluntarily assuming additional responsibility in the digital context, especially with regard to data privacy and security, is becoming increasingly important for companies and, above all, making this behavior visible to consumers – usually by communicating such activities. The topic of data privacy and security is multifaceted and does not occur isolated from further responsibilities in the digital era. Accordingly, a concept such as CDR can help to put these responsibilities into context and thus track consistent communication regarding digital responsibilities. This also favors the communication of data privacy and security related activities and creates a certain credibility and consumers' trust. Within the context of CDR, ethically responsible behavior and communication related to data privacy and security forms one dimension of

the concept [e.g., Th17, Mi21] and comprises various fields of action. Several different classification approaches developed in parallel. In order to keep the complexity bearable for practitioners, this study employs a classification based on the four privacy areas according to Smith et al. [SMB96] and current legislation (e.g., GDPR, OECD guidelines [Oe13]). Table 1 summarizes the data privacy and security fields of action according to the concept of CDR.

Fields of action	Description	Related work
Data privacy and security fields of action		
Openness	Openness describes the degree to which companies are open with consumers, especially with regard to data processing practices.	GCC15, THS08
Data collection	Data collection covers limited data collection and the clear purpose of data collection.	GCC15, THS08
Data processing	Data processing comprises restricted data use as well as secure storage and processing of data.	BC11, GCC15
Data management	Data management refers to data quality of user data (e.g., accurate, up-to-date) and consumers' access and correction opportunities.	SK08, SMB96
Further data privacy and security related fields of action within the concept of CDR		
Transparency on internal governances	Transparency on internal data privacy and security governances describes the disclosure of internal practices to consumers.	ABK17, MW10
Transparent communication on internal practices	A transparent communication on data privacy and security addresses additional information on data transmission practices and a transparent and easy to understand data protection declaration.	BC11, GS09
Access without data input	Access without data input describes consumers' ability to access products and services (e.g., advisory services) without entering personal data.	BC11, SK08

Tab. 1: Data privacy and security fields of action within the concept of CDR

3 Study Design and Sample

To derive best practices on ethically responsible data privacy and security communication we rely on qualitative text analysis. Therefore, we quota sampled on companies operating in Germany from different industries and with varying digital products and services. The sample is limited to companies operating in Germany due to the already high legal and regulatory requirements for corporate communication on data privacy and security and the increased awareness of consumers. With the inclusion of varying products and industries, this study aims at creating a cross-industry understanding of ethically responsible

corporate communication in this field. To derive best practices and therefore guidance, we rely on the communication of large companies as they usually have a larger budget available for possible activities and their communication. Besides, larger companies primarily participate in initiatives for more ethical behavior. When looking at industry initiatives (e.g., the Bundesverband Digitale Wirtschaft e.V.) as well as governmental initiatives (e.g., the CDR Initiative of the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety, and Consumer Protection), it is mainly larger companies that participate to set new industry standards. Hence, we included companies with such a visible CDR commitment primarily in the sample, as their motivation is the setting of best practices. Accordingly, they can mostly be used for such, especially at such an early stage of CDR implementation in practice. Hence, we selected a sample of 26 companies (see Table 2) based on the targeted criteria. We crawled potentially relevant corporate communication from the website and qualitatively analyzed them. The documents range from legal documents (e.g., data protection declarations) to voluntary communication of corporate responsibility. Only publicly available documents were included in the sample[2], as only these are accessible to consumers. Therefore, this study excludes company internal documents not available for people outside the company. We collected the data in fall 2021 to draw a recent picture of corporate communication on data privacy and security.

Industry[3]	**Company**
Automotive	BMW Group, Continental AG, Mercedes-Benz Group, Volkswagen AG
Banking and financial services	HSBC Trinkaus & Burkhardt AG, ING Bank N. V., SCHUFA Holding AG
Computer hardware & software	IBM Corporation, Microsoft Corporation
Cosmetics & pharmacy	L'Oréal Groupe, Merck KGaA
Electrical appliances	Miele & Cie. KG, Philips GmbH Market DACH, Robert Bosch GmbH, Siemens AG, Sony Group, Vorwerk SE & Co. KG
Insurance	BARMER, Zurich Insurance Group Ltd
Retail and consumer services	Otto Group, Rewe Group, Zalando SE
Telecommunication	Deutsche Telekom AG, Telefónica S.A., Vodafone Group Plc
Utilities	EWE AG

Tab. 2: Companies in the study sample

We used MAXQDA for the qualitative analysis of corporate communication. No traditional coding scheme matches the topic under investigation. Hence, we developed a coding scheme according to established guidelines for thematic qualitative text analysis [Ku14, pp. 69ff.] and accordingly developed our codebook during the coding process.

[2] The study analyzed 766 documents in total, thus in average 29 documents per company.
[3] Assignment based on industry affiliation usually known to the consumer.

Three independent researchers performed the coding in parallel to discuss and contrast coding results. This ensures an adequate quality of results [VBS16]. The team reviewed conflicts in coding and solved them in mutual agreement, thus systematically developing and adapting the codebook. An external coder with prior coding experience assessed the coding results to ensure semantic quality and agreed with the assignments.

4 Results

Our findings indicate that most information on data privacy and security related activities is of a legal nature and less about the company assuming additional responsibility towards the consumer. However, for example in the EU, the websites under consideration are also required to publish (legal) privacy statements. Nevertheless, in the course of assuming additional responsibility on a voluntary basis, companies can provide additional documents and website content that exceeds these legal requirements. Addressing these additional responsibilities, this work-in-progress takes a closer look at possible best practices for data privacy and security related corporate communication under the concept of CDR. The aim of the study is to identify some examples of additionally assumed responsibility per data privacy and security field of action and thus to provide guidance for researchers and practitioners alike on responsible corporate communication. However, this study does not intend to provide an all-encompassing catalogue of possible best practices or to provide a comparison between the investigated companies. Accordingly, the frequency of the companies cited should not be used to evaluate a company's corporate communication. Rather, each field of action comprises various possible corporate communication examples that represent best practices. In the following, exemplary activities in corporate communication are illustrated for the various fields of action and best practices are discussed. Communication best practices should also serve as guidelines for ethical behavior, ensuring that communication and internal behavior are aligned.

The field of action related to openness describes responsible acting of companies by disclosing their data processing practices openly in the course of additional information [e.g., GCC15, THS08]. For instance, companies can provide additional website content addressing further, more comprehensible openness about data protection practices also disclosing precisely what the data is used for: "*For example, insights from the data will enable Zurich to provide innovative services that help to prevent incidents, expanding the traditional protection offered by insurance. These include smart services for home protection, and to improve health and well-being, as well as travel that keeps customers out of harm's way.*" [Zu21]. In a similar vein, companies can aggregate the main purposes for which consumer data is used, thus demonstrating greater openness to consumers: "*Key uses of customer data are outlined below. […]*" [Vod21]. Such openness supports companies' responsible communication towards consumers about their data processing practices and thus informed consumer decision making.

Data collection concerns limited or restricted data collection on the one hand and the clear

purpose of data collection on the other hand [e.g., GCC15, THS08]. For example, companies can address this additionally assumed responsibility by providing additional website content explaining limited data collection practices: "*[W]e gather only the data that is absolutely essential, […] as well as ensuring transparency vis-à-vis the customer.*" [De21]. Besides, companies can explain and communicate their default settings of products and services regarding (limited) data collection: "*[A]ppliances are shipped out with default settings that ensure the highest degree of data minimisation. Any communication or data transfer that goes beyond this must be explicitly activated by the customer.*" [Mie21]. Especially the application of more advanced systems, like AI-based systems, deepen the necessity of taking responsibility with regard to data collection. For instance, companies should therefore address these consumer concerns, disclosing their internal obligation for limited and purposeful data collection: "*The aggregation and use of customer data – especially in AI systems – shall always be clear and serve a useful purpose towards our customers.*" [De21]. Corporate communication geared to these demands are evidence of more ethical corporate behavior.

Data processing practices address manifold facets of responsible behavior including handling of secondary usage, transfer of data to third parties, or secure processing and storage of data, therefore presenting another important field of action [e.g., BC11, GCC15]. Thus, companies can assume additional responsibilities in manifold ways, e.g., offering information on data processing to third parties as additional website content: "*Further, any third party with whom Zurich does share personal data is bound by an enforceable contract, which sets out how that personal data can be used.*" [Zu21, see also Ph21]. Besides, additional information and even possible ways of contacting regarding secure storage and processing of data is another path to assume more responsibility regarding data processing: "*But we are extremely conscious of our responsibilities to all our stakeholders and work diligently to secure business assets and information, especially data protected under the law and sensitive information under antitrust law. […] We always comply with the processes on data and information security. We expect you to ask for advice with the Privacy & Technology Law team (it-law-intern@zalando.de) whenever you are in doubt.*" [Za21, see also Ph21]. Another approach is to inform in detail about data processing and securing mechanisms like the privacy by design process applied internally to product and service development [Te21]. Besides, companies can provide additional information on certifications or passed security tests: "*In this context, we at Vorwerk are proud to be the winners of independent tests on data security. For example, most recently in January 2019, the Kobold VR300 Vacuum Robot won the security check undertaken by the independent IT security institute AV-TEST.1.*" [Vor21, see also Ot21]. Hence, due to the many different aspects of responsible data processing, there are also many different approaches for companies to communicate ethical behavior.

Another field of action, precisely data management, covers data quality as well as consumers' access to their stored data [e.g., SK08, SMB96]. Therefore, companies acting ethically responsible can pursue and communicate to consumers that they "*[m]anage data carefully*" including to "*[m]aintain data quality, delete unnecessary or outdated data and do not take unnecessary copies.*" [Vod21, see also Hs21, In21, Sc21]. Besides, companies

should facilitate and communicate consumers' access and deletion rights to their data stored: "*[We d]esign products to allow individuals to exercise their right to access, request deletion and portability.*" [Vod21]. This approach also includes to "*[e]mpower our customers through simple and secure tools so that they can control the use of their personal data.*" [Te21, see also In21, Ph21, Ro21], therefore respecting individual rights related to stored data.

Besides, companies can voluntarily provide transparency regarding internal governances affecting data privacy and security activities [e.g., ABK17, MW10]. For instance, companies can disclose internal governances directly addressing their data privacy and security related activities, e.g., "*[we] have created, among other things, a uniform international framework for this in the shape of our Binding Corporate Rules Privacy. These define the purposes for which personal data may be collected, stored and processed. We endeavor to clarify any unresolved questions or issues without delay. We set ourselves principles early on for new business models, which make it clear to our customers, employees and business partners how we deal with these models and leave them in no doubt that we assume responsibility.*" [De21] or disclosing additional information on their risk management principles related to data privacy and security: "*The Zurich Risk Policy, applied across the Group, lays down the Group´s risk management principles. There is a dedicated section on Information Security supported by detailed policy manuals, guidelines and standards including those for data handling and classification of assets and information.*" [Zu21, see also Ph21] accompanied by associated training material [Ew21, Sc21, Si21, Vor21]. Besides, companies also establish own codes of ethics to ensure ethical behavior in general [Ba21, Lor21, MK21, Re21, Ro21, Si21, Za21], specific AI guidelines for ethical AI deployment and usage [Bm21, Co21, De21, Ib21, Mic21, Re21, Ro21, So21], or CDR guidelines [Sc21]. Internal governances also include the disclosure of specific boards [MK21] or committees [Ph21] established to foster additional data privacy and security activities: "*We particularly focus on maintaining information security and defending against cyber attacks. The Group Information Security Steering Committee (GISSC) [..] has overarching responsibility.*" [Vol21]. By disclosing such internal governances, practices, and initiatives, companies can prove ethically responsible behavior and transparently communicate it to consumers.

Earlier research has already addressed the need for more understandable privacy statements for a considerable time [e.g., BC11, GS09]. Nevertheless, there has not been that much progress yet in this field and many companies leave it primarily to the publication of the legally binding privacy policies. This is reflected above all in relatively long privacy statements, which are also worded in a rather complicated manner and thus do not allow consumers to easily understand the data distribution, collection, and management practices. However, some companies already make some effort to provide easier to understand additional documents. For example, companies can provide a "*"one-pager": an easy-to-read, brief overview of the main data processing activities*" [De21, see also MB21]. Such a one-pager facilitates consumers' access to relevant data privacy and security activities of a company and thus provides a reasonable approach to additional assumption of responsibility in the digital context. In addition, there are discussions in

public about using data protection icons to ensure easier access to data protection declarations. Yet, such an approach could not be observed in the sample studied.

Another field of action concerns access to digital services or products without data input [e.g., BC11, SK08]. Companies can offer access to several digital products and services without requiring the disclosure of personal data. This additional responsibility is applicable to a wide range of services and products and also includes, e.g., the topping up of credit without disclosing personal data: "*Provides a dummy ten-digit number to ensure the privacy of customers when they recharge at retail outlets, avoiding the need for them to have to reveal their mobile number to an unknown retailer.*" [Vod21]. Activities and corporate communication related to this field of action can reduce consumers' privacy concerns, as disclosure of personal data is not forced.

5 Conclusion and Outlook

As results illustrate, we can already find a vast set of best practices regarding data privacy and security corporate communication. Therefore, it is of tremendous importance for companies of all sizes and industries to engage in the assumption of additional digital responsibilities and to address this topic in the future. This work-in-progress can serve as a first starting point for future paths and inspiration for ethically responsible corporate communication related to data privacy and security. Best practices could be identified in various data privacy and security fields of action that can guide future paths of corporate communication. By providing an ethical evaluation of the status-quo of corporate communication related to data privacy and security, this study benefits researchers and practitioners alike. Theoretically, this work in progress supports the understanding of ethically responsible corporate communication on data privacy and security. Besides, this study emphasizes the comprehensive scope of data privacy and security related activities when viewed in the larger context of concepts such as CDR. Practically, this study serves as a starting point for various companies with a B2C focus for a more ethically responsible corporate communication and therefore future paths for implementing additional activities or communicating already performed endeavors. However, ethically responsible communication does not always mean ethical behavior. While corporate communication can be ethically responsible, a company's behavior can still be unethical. Besides, companies (often) focus more on their own needs than on the ethically responsible treatment of consumer interests. Accordingly, this work aims to promote not only ethically responsible communication but also appropriate behavior and to reinforce the importance of it.

Despite best efforts, this study is not without limitations. The sample is limited to companies operating in Germany, especially B2C companies, and their communication addressing consumer concerns. Hence, future research should add to the derived best practices and future paths for data privacy and security communication by including companies from different countries. Besides, B2B companies and other stakeholders of

companies (e.g., employees) should complement a broader understanding of best practices in corporate communication. Nevertheless, this study can provide a first starting point for theory and practice for a more ethically responsible data privacy and security behavior and its communication, thus providing first guidance on the this topic.

Acknowledgement

The Hessian State Chancellery – Hessian Minister of Digital Strategy and Development supported this work under the promotional reference 6/493/71574093 (CDR-CAT).

Bibliography

[ABK17] Anderson, C.; Baskerville, R. L.; Kaul, M.: Information Security Control Theory: Achieving a Sustainable Reconciliation Between Sharing and Protecting the Privacy of Information. Journal of Management Information Systems 34/4, pp. 1082-1112, 2017.

[Ba19] Baumann, A.; Haupt, J.; Gebert, F.; Lessmann, S.: The Price of Privacy: An Evaluation of the Economic Value of Collecting Clickstream Data. Business & Information Systems Engineering 61/4, pp. 413-431, 2019.

[Ba21] BARMER, https://www.barmer.de/en, accessed: 29/10/2021.

[BC11] Bélanger, F.; Crossler, R. E.: Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. MIS Quarterly 35/4, pp. 1017-1041, 2011.

[Bm21] BMW Group, https://www.bmwgroup.com/en.html, accessed: 29/10/2021.

[Co21] Continental AG, https://www.continental.com/en/, accessed: 29/10/2021.

[CZH22] Carl, K. V.; Zilcher, T. M. C.; Hinz, O.: Corporate Digital Responsibility and the Current Corporate Social Responsibility Standard: An Analysis of Applicability. In: Proceedings of the Open Identity Summit 2022 (OID2022), 2022.

[De21] Deutsche Telekom AG, https://www.telekom.com/en, accessed: 29/10/2021.

[Ew21] EWE AG, https://www.ewe.com/, accessed: 29/10/2021.

[GCC15] Greenaway, K. E.; Chan, Y. E.; Crossler, R. E.: Company Information Privacy Orientation: A Conceptual Framework. Information Systems Journal 25/6, pp. 579-606, 2015.

[GS09] Goel, S.; Shawky, H. A.: Estimating the Market Impact of Security Breach Announcements on Firm Values. Information & Management 46/7, pp. 404-410, 2009.

[Ha07] Hann, I.-H.; Hui, K.-L.; Lee, S.-Y. T.; Png, I. P. L.: Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach. Journal of Management Information Systems 24/2, pp. 13-42, 2007.

[He21] Herden, C.; Alliu, E.; Cakici, A.; Cormier, T.; Deguelle, C.; Gambhir, S.; Griffiths, C.; Gupta, S.; et al.: “Corporate Digital Responsibility”: New Corporate Responsibilities in

the Digital Age. Sustainability Management Forum 29, pp. 13-29, 2021.

[HH18] Heimbach, I.; Hinz, O.: The Impact of Sharing Mechanism Design on Content Sharing in Online Social Networks. Information Systems Research 29/3, pp. 592-611, 2018.

[Hs21] HSBC Trinkaus & Burkhardt AG, https://www.about.hsbc.de/, accessed: 29/10/2021.

[Ib21] IBM Corporation, https://www.ibm.com/us-en?lnk=fcc, accessed: 29/10/2021.

[In21] ING Bank N. V., https://www.ing.com/Home.htm, accessed: 29/10/2021.

[IP18] Intezari, A.; Pauleen, D. J.: Conceptualizing Wise Management Decision-Making: A Grounded Theory Approach: Conceptualizing Wise Management Decision-Making. Decision Sciences 49/2, pp. 335-400, 2018.

[Ku14] Kuckartz, U.: Qualitative Text Analysis: A Guide to Methods, Practice & Using Software. Sage Publications, London & Thousand Oaks, 2014.

[Lob21] Lobschat, L.; Mueller, B.; Eggers, F.; Brandimarte, L.; Diefenbach, S.; Kroschke, M.; Wirtz, J.: Corporate Digital Responsibility. Journal of Business Research 122, pp. 875-888, 2021.

[Lor21] L'Oréal Groupe, https://www.loreal.com/en/, accessed: 29/10/2021.

[Lu02] Luo, X.: Trust Production and Privacy Concerns on the Internet: A Framework Based on Relationship Marketing and Social Exchange Theory. Industrial Marketing Management 31/2, pp. 111-118, 2002.

[Ma86] Mason, R. O.: Four Ethical Issues of the Information Age. MIS Quarterly 10/1, pp. 5-12, 1986.

[MB21] Mercedes-Benz Group, https://group.mercedes-benz.com/en/, accessed: 29/10/2021.

[Mic21] Microsoft Corporation, https://www.microsoft.com/en-us/, accessed: 29/10/2021.

[Mie21] Miele & Cie. KG, https://www.miele.com/en/com/index.htm, accessed: 29/10/2021.

[Mi21] Mihale-Wilson, A. C.; Zibuschka, J.; Carl; K. V.; Hinz, O.: Corporate Digital Responsibility – Extended Conceptualization and a Guide to Implementation. In: Proceedings of the European Conference on Information Systems (ECIS) 2021, 2021.

[Mi22] Mihale-Wilson, A. C.; Hinz, O.; van der Aalst, W.; Weinhardt, C.: Corporate Digital Responsibility: Relevance and Opportunities for Business and Information Systems Engineering. Business & Information Systems Engineering 64/2, pp. 127-132, 2022.

[MK21] Merck KGaA, https://www.merckgroup.com/en, accessed: 29/10/2021.

[MR02] Maignan, I.; Ralston, D. A.: Corporate Social Responsibility in Europe and the U.S.: Insights from Businesses' Self-Presentations. Journal of International Business Studies 33/3, pp. 497-514, 2002.

[MW10] Mingers, J.; Walsham, G.: Toward Ethical Information Systems: The Contribution of Discourse Ethics. MIS Quarterly 34/4, pp. 833-854, 2010.

[MZH17] Mihale-Wilson, A. C.; Zibuschka, J.; Hinz, O.: About User Preferences and Willingness to Pay for a Secure and Privacy Protective Ubiquitous Personal Assistant. In: Proceedings of the European Conference on Information Systems (ECIS) 2017,

pp. 32-47, 2017.

[Oe13] OECD, http://www.oecd.org/sti/ieconomy/oecd_privacy_framework.pdf, accessed: 09/07/2019.

[Ot21] Otto Group, https://www.ottogroup.com/en/, accessed: 29/10/2021.

[Ph21] Philips GmbH Market DACH, https://www.philips.com/global, accessed: 29/10/2021.

[Re21] Rewe Group, https://www.rewe-group.com/en/, accessed: 29/10/2021.

[Ro21] Robert Bosch GmbH, https://www.bosch.com/, accessed: 29/10/2021.

[Sc21] SCHUFA Holding AG, https://www.schufa.de/schufa-en/, accessed: 29/10/2021.

[Si21] Siemens AG, https://www.siemens.com/global/en.html, accessed: 29/10/2021.

[SK08] Son, J.-Y.; Kim, S. S.: Internet Users' Information Privacy-Protective Responses: A Taxonomy and a Nomological Model. MIS Quarterly 32/3, pp. 503-529, 2008.

[SMB96] Smith, H. J.; Milberg, S. J.; Burke, S. J.: Information Privacy: Measuring Individuals' Concerns About Organizational Practices. MIS Quarterly 20/2, pp. 167-196, 1996.

[So21] Sony Group, https://www.sony.com/en/, accessed: 29/10/2021.

[Te21] Telefónica S.A., https://www.telefonica.com/en/, accessed: 29/10/2021.

[Th17] Thorun, C.; Vetter, M.; Reisch, L.; Zimmer, A. K., https://www.bmjv.de/G20/DE/ConsumerSummit/_documents/Downloads/Studie.pdf?__blob=publicationFile&v=1, accessed: 09/07/2019.

[THS08] Tang, Z.; Hu, Y.; Smith, M. D.: Gaining Trust Through Online Privacy Protection: Self-Regulation, Mandatory Standards, or Caveat Emptor. Journal of Management Information Systems 24/4, pp. 153-173, 2008.

[VBS16] Venkatesh, V.; Brown, S.; Sullivan, Y.: Guidelines for Conducting Mixed-methods Research: An Extension and Illustration. Journal of the Association for Information Systems 17/7, pp. 435-494, 2016.

[Vod21] Vodafone Group Plc, https://www.vodafone.com/, accessed: 29/10/2021.

[Vol21] Volkswagen AG, https://www.volkswagenag.com/en.html, accessed: 29/10/2021.

[Vor21] Vorwerk SE & Co. KG, https://www.vorwerk-group.com/en/home, accessed: 29/10/2021.

[We67] Westin, A. F.: Privacy and Freedom. 1st. ed. Athenum, New York, 1967.

[Za21] Zalando SE, https://corporate.zalando.com/en, accessed: 29/10/2021.

[Zu21] Zurich Insurance Group Ltd, https://www.zurich.com/en, accessed: 29/10/2021.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 207

Authentication and Authorization in Microservice-Based Applications

Niklas Sänger[1], Sebastian Abeck[2]

Abstract: The development of microservice-based applications adds challenges when using different cloud services. One such challenge is the integration of authentication and authorization among different systems. In this publication, we describe the development of a software as a service solution with the focus on the integration of authentication and authorization. For the development of the business logic, the integration platform as a service MuleSoft is used. The identity and access management as a service solution Okta is used to provide the necessary means for authentication. To perform authorization decisions, JSON Web Tokens and API proxies are used.

Keywords: Authentication; Authorization; Microservices; Engineering; Okta; MuleSoft

1 Introduction

Compared to the development of a classical monolithic application, the goal of a modern cloud application using a microservice architecture is to reuse existing software components or services by integrating them through well-defined application programming interfaces (APIs) [Ne15]. This publication describes the integration of authentication and authorization in the development of a Software as a Service (SaaS) using the products MuleSoft and Okta.

As a demonstrator, the case study of the development of a connected car application, called PredictiveCarMaintenance (PCM), is used. Fig. 1 provides an overview of the groups accessing as well as the products and protocols used to build the application PCM which provides vehicle data (e.g., state, maintenance intervals) for vehicle users and garages. For the implementation of the application, the well-established products Okta and MuleSoft are used. Okta provides an IAM as a Service (IAMaaS) solution to, among others, register and authenticate users [In22b]. MuleSoft calls itself an Integration Platform as a Service (iPaaS), which can be used to develop tailored APIs by integrating existing systems [Mu22b]. In order to meet the IAM requirements, the two products must be strictly aligned with each other. The established protocols OpenID Connect (OIDC) and OAuth 2.0 are used for this purpose. To allow a user to login and display the requested data, the User Interface (UI) is developed using the TypeScript-based framework Angular.

[1] Karlsruhe Institute of Technology, Research Group Cooperation & Management, Zirkel 2, 76131 Karlsruhe, Germany, niklas.saenger@kit.edu
[2] Karlsruhe Institute of Technology, Research Group Cooperation & Management, Zirkel 2, 76131 Karlsruhe, Germany, sebastian.abeck@kit.edu

doi:10.18420/inf2022_19

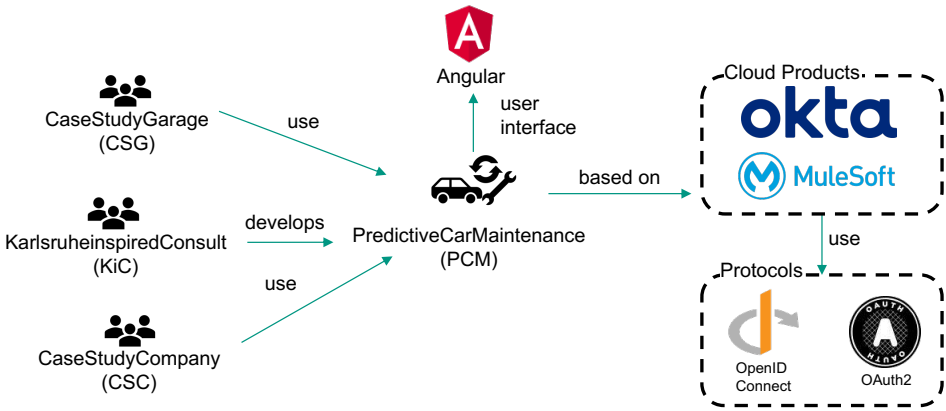


Fig. 1: Overview of the Application PredictiveCarMaintenance

Following the principle of least privilege, the users of the application PCM should only have the privileges to perform their necessary tasks [SS75]. Those tasks differ for each group and should be defined by a detailed requirements analysis. In the context of this publication, the groups, their tasks, and the requirements are described informally in the following.

PCM is developed by the company KarlsruheinspiredConsult (KiC). The CaseStudyCompany (CSC) is a customer of KiC which has a large vehicle fleet and uses the application PCM to manage their fleet. Finally, the company CaseStudyGarage (CSG) is a garage and uses the application PCM to organize vehicle repairs. Since employees from KiC are the developers of the application PCM, they must also perform administrative tasks as well as support for their customers. Therefore, KiC employees require access to all data that is available in the system. To protect these comprehensive permissions, a second factor is required for authentication of KiC employees. This factor must be verified by the IAMaaS Okta. Employees of CSC use the application PCM to see current and past data of the vehicles they are assigned to. An employee is assigned to a vehicle by the company CSC. While an employee can schedule a maintenance appointment with the CSG through the application PCM, the CSC employee cannot access the advanced diagnostics data and maintenance information. Vice versa, employees of the CSG must only access the diagnostics data and maintenance information of a vehicle they have been assigned to by a supervisor.

The present article is structured as follows: Sect. 2 presents the related work. The design of the frontend architecture and the implementation of the authentication based on the previously described requirements is described in Sect. 3. The design and the implementation of the authorization are described in Sect. 4. Finally, the results and future work are discussed in Sect. 5.

2 Related Work

In their systematic review from 2012, Iankoulova and Davena analyze security requirements for cloud computing and find that access control, as the degree to which a system limits access to its resources, is a key requirement [ID12]. This is also a problem in a microservice architecture as de Almeida and Canedo [AC22] point out in their systematic literature review on authentication and authorization in microservice architectures. They find that challenges are trust between microservices compromised by unauthorized access, individual concern for each microservice, and microservice access control. Furthermore, the authors find the most used security mechanisms are OAuth 2.0, JSON Web Tokens (JWT), and API Gateways. Other important mechanisms are Single-Sign On, OpenID Connect (OIDC), as well as Role Based Access Control (RBAC) and Attribute Based Access Control (ABAC).

The Open Web Application Security Project (OWASP) further strengthens the security issues in their report of the top 10 security risks in web applications [Fo21]. The 2021 report includes broken access control (A01:2021) as the most serious web application security risk.

He and Yang [HY17] describe the realization of authentication and authorization of an end-user in a microservice architecture. They propose an authorization solution using JWTs provided by an authentication server and an API gateway to validate the JWT. Zhao et al. [ZJJ18] also describe an authorization mechanism using an API gateway and JWT issued by an OAuth 2.0 authorization server. Xu et al. introduce the development of a microservice security agent in an edge computing environment [XJK19]. The authors propose a security agent consisting of an API gateway, a client server including a user interface, and a security agent. The client server is used to configure the API gateway. The API gateway then performs authorization decisions based on a JWT which are issued by the security agent. Similar to the described approaches, this publication performs authentication and authorization decisions based on an API gateway as well as JWTs. However, this publication sets the focus on the realization based on the integration of the cloud solutions MuleSoft and Okta.

Nehme et al. describe fine-grained authorization for a microservice architecture based on OAuth 2.0 and eXtensible Access Control Markup Language (XACML) [Ne19]. They develop a gateway solution which applies XACML policies to resource requests, thus decoupling the security aspects from functional requirements of a microservice. Furthermore, the authors describe the shortcomings of the coarse-grained nature of OAuth 2.0 which is used in this publication.

MuleSoft offers the API-led connectivity approach, which aims at connecting data to applications through reusable APIs [Mu22a]. To do so, MuleSoft uses a three-layered architecture with each layer having its own API type. First, the System layer is the lowest layer and provides System APIs integrating the data from external systems. The middle layer is the Process layer which performs business logic and provides the results through Process APIs. Finally, the top layer is the Experience layer providing the processed data

in the required format for an arbitrary client through an Experience API. This layer can be compared to a Backend for Frontend (BFF), which does not perform business logic but rather supports the frontend [Ne15].

3 Authentication

This chapter describes the systematic realization of the authentication requirements for the application PCM described in Sect. 1. To restrict access to the application PCM for authenticated users, the architecture of the application PCM must take authentication into account during the design phase. Next, the implementation phase consists of two steps. First, the configuration of the IAM system Okta to implement the required authentication mechanisms (i.e., OIDC Client, multifactor authentication). Second, the implementation of the frontend using the results of the Okta configuration.

3.1 Design of the Architecture

The UML-based diagram displayed in Fig. 2 presents an excerpt of the architecture of the application PCM with a focus on the frontend. It is a combination of a UML deployment diagram and a UML component diagram [OM17]. It allows to further enrich a physical node from a deployment diagram by applying software components from a UML component diagram running on the subsystems. It also presents a more detailed view on which and how components inside physical nodes communicate with each other (i.e., protocol).

Based on the case study requirements depicted in Sect. 1, the diagram in Fig. 2 contains three nodes for the involved systems «iamaas system» Okta and «ipaas system» MuleSoft as well as «browser» PCMFrontend for the required Angular frontend. While the architecture of the frontend is generally independent from the implementation, the components are enriched by Angular-specific stereotypes «ng component» and «ng service» to differentiate views from reusable functions or tasks (e.g., authentication).

To allow a user to log into the PCMFrontend, the «ng component» Login provides a web page with fields for a username and password as well as a login button. The «ng service» Authentication, initiates the authentication by executing the OIDC flow against the OAuth «authorization server» Default provided by Okta. The «ng service» Authentication uses the «oidc client» PredictiveCarMaintenanceOpenidconnectClient (PCMOC) to authorize itself to the «authorization server» Default. After a user is successfully authenticated, they are redirected to the «ng component» LandingPage which provides references to the pages a user is allowed to access. The design of the components required for authentication and authorization is inspired by the OIDC flow and allows for separation of concerns in the frontend architecture.

Furthermore, for each user group (i.e., VehicleUser, Garage), a dedicated «ng component» is created to display the requested information. To retrieve the information, an experience

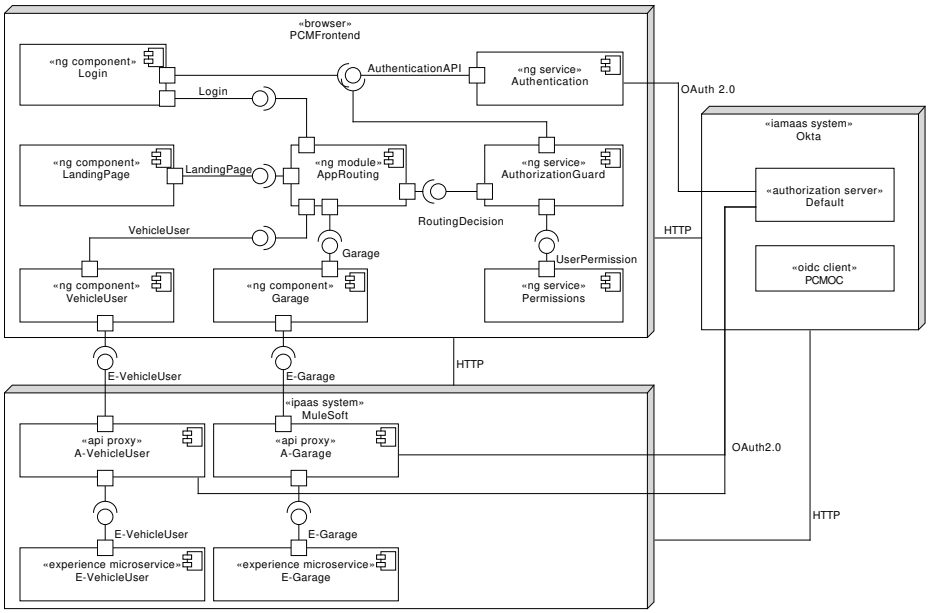


Fig. 2: Excerpt of the PCMFrontend with Focus on Authentication and Authorization

microservice is developed for each user group. This allows to only provide the data required for a user group and further restrict the access to the experience microservice with the use of an API proxy. This is also inspired by the MuleSoft API-led connectivity approach [Mu22a]. The authorization decisions in the «browser» PCMFrontend and the experience microservices are further elaborated in Sect. 4.

3.2 Configuration of the IAMaaS Okta

Before users can authenticate themselves in the PCMFrontend and access sensitive data, the IAM system Okta must be configured accordingly. Therefore, the first thing that must be decided is where and how the users will be stored inside Okta. In this example, the users are created manually inside Okta. Okta offers a universal directory which provides a single view on users and groups and provides integration of further directories (e.g., LDAP, Active Directory). To be able to distinguish between the affiliation and the function of a user (e.g., a mechanic or vehicle user), each user is assigned to a corresponding group which must first be created in Okta. For PCM, the groups CaseStudyCompany, CaseStudyGarage, and KarlsruheinspiredConsult are created for the employees of the respective companies being part of the case study.

The OIDC protocol is based on OAuth 2.0 and allows authenticating users to an identity provider. In this example, the IAM system Okta acts as the identity provider and provides an OIDC client to the PCMFrontend. Depending on the requirements, there are several options to configure and use an OIDC client (e.g., client types, grant types). Okta guides the developer through the creation of an OIDC client by asking for the application which has implications on the basic configuration. In case of PCMOC, the SPA application type is used. This forces the developer to use a public OIDC client since the client ID and secret can be extracted from the deployed frontend otherwise. Furthermore, the developer has to use Proof Key for Code Exchange (PKCE) which is also recommended by the Internet Engineering Task Force (IETF) for public OAuth 2.0 clients [SBA15]. To be able to use the OIDC client and authenticate the users to the PCMFrontend, valid redirect URLs must be configured (e.g., *https://localhost:4200/home*). The redirect URLs are validated during the authentication to ensure that a request is coming from the PCMFrontend.

So far, the users of the group KarlsruheinspiredConsult can access the PCMFrontend through OIDC client PCMOC with their username and password. However, since the KiC employees have vast access to the application PCM, the authentication is further strengthened by introducing a second factor. This process is called Multi-factor Authentication (MFA). Okta can enable MFA per OIDC client or per Okta organization. It is also possible to enable the MFA for selected users or groups. The configuration of MFA is done through an authentication policy. For the application PCM, time-based one-time-passwords (TOTP) are selected as a second authentication factor [Vi11].

To enforce the second factor, a new multifactor authentication policy is created and enrolled for the group KarlsruheinspiredConsult. Enabling the MFA policy changes the authentication process displayed in Fig. 3. After the user is prompted the authentication form and enters the credentials, the authorization server verifies the credentials and checks if there is an MFA policy enabled for the user. If the user is from KiC, the user is asked to enter a TOTP. After the user enters the correct one-time password, the authentication process is continued and the user is successfully authenticated using MFA. Otherwise, the authentication process fails, and the user is not authenticated.

3.3 Implementation of the Authentication

For the implementation of the authentication, Okta allows to either sign in users through a redirect to their page or to use an embedded widget which performs the authentication in the background. The PCMFrontend uses the customized embedded Okta sign-in widget to provide users a seamless interface throughout the authentication process [In22a].

The authentication process for a user is further described in the UML sequence diagram in Fig. 3. The goal of the authentication process is to obtain an ID token as well as an access token. After a user opens the «ng component» Login, the «ng service» Authentication is called, which uses the Okta sign-in widget to initiate the OIDC flow with PKCE. Therefore,

the «ng service» Authentication first has to generate a verifier as well as a code challenge. The verifier is a random string with a length between 43 and 128 characters. The code challenge is the SHA256 hashed value of the verifier [SBA15]. The code challenge is then included in an HTTP GET request to the endpoint */authorize* of the OAuth 2.0 «authorization server» Default. Afterwards, the «ng component» Login presents the embedded Okta sign-in widget. The user can then enter their credentials (i.e., username and password) and press the login button. The credentials are then parsed to the «authorization server» Default and are subsequently verified. Otherwise, an authorization code is returned to the «ng service» Authentication. The authorization code can then be exchanged by the «ng service» Authentication to obtain the requested tokens. Therefore, the «ng service» Authentication has to send an HTTP POST request which includes the authorization code as well as the code verifier to the */token* endpoint of the «authorization server» Default. The «authorization server» Default can then verify the authorization code. To verify the code verifier, the SHA256 hash function must first be applied to the code verifier and the result compared to the code challenge. If the authorization code and code verifier can be verified, the ID token and the access token are returned to the «ng service» Authentication. Finally, the user is redirected to the «ng component» LandingPage.

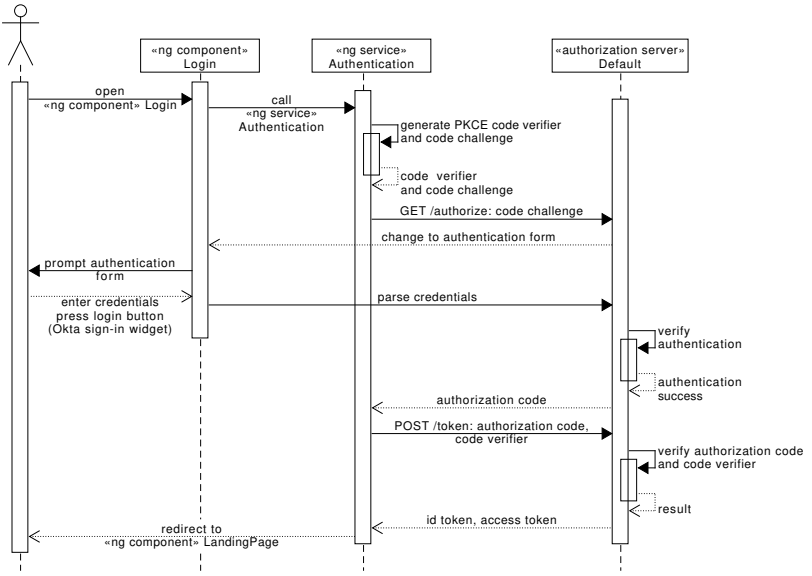


Fig. 3: Sequence Diagram of the Authentication Flow

4 Authorization

The realization of authorization is required to protect the user data of the application PCM. Similar to the authentication, the authorization is an essential aspect of the design phase and consequently realized in the implementation phase. To perform authorization decisions, this section will describe the design of the access control mechanisms in the frontend and backend. For the application PCM, the access control decision has to be implemented in two places. First, the PCMFrontend has to decide what view a user can access based on the group membership of a user (i.e., Attribute Based Access Control). Second, the iPaaS system MuleSoft must decide what entities a user can access on an API-level.

4.1 PCMFrontend-level Authorization

After the successful authentication, the access token is returned from the authorization server Default to the PCMFrontend. Listing 1 presents an excerpt of an access token provided by Okta. The access token contains requested information in the JWT format and is used by the PCMFrontend to decide which «ng component» a user can access [JBS15].

In the architecture of the PCMFrontend presented in Fig. 2, the «ng service» Authorization Guard and the «ng service» Permissions are used to perform the authorization. The «ng service» Permissions contains a mapping from group names to paths (i.e., «ng component») that a user is allowed to access. The AuthorizationGuard examines the access token and extracts the claim groups. In the example access token of Listing 1, the user is in the groups Everyone and CaseStudyCompany (line 5). Based on the extracted groups, the AuthorizationGuard decides if the user can access a path or not. If a user only has the group CaseStudyCompany in their access token, the access to the «ng component» VehicleUser is allowed, while the access to the «ng component» Garage is denied.

```
1.  "exp": 1643886904,
2.  "cid": "0oa3ouku57Q89YCEc0x7",
3.  "scp": [ "openid", "email", "profile" ],
4.  "sub": "alex.twin@csc.com",
5.  "groups": [ "Everyone", "CaseStudyCompany" ]
```

List. 1: Excerpt of an Access Token Provided by Okta

4.2 API-Level Authorization

Protecting the access to an «ng component» is not enough since the «ng component» obtain their data from a RESTful API which could otherwise be directly accessed. Thus, the access must also be protected on an API-level. There are two challenges that generally

must be overcome: First, coarse-grained authorization requires to protect the API itself (i.e., who can access a specific API endpoint). Second, fine-grained authorization decides what resources a user can access and which operations they can perform. A common approach to solve coarse-grained authorization, which is also supported by MuleSoft, is the use of API gateways as a mediator between a client and a service. An API gateway allows to create API proxies and apply and enforce policies to such a proxy. As depicted in Fig. 2, the application PCM has the «experience microservice» E-VehicleUser and the «experience microservice» E-Garage which provide the data for the «ng component» VehicleUser and the «ng component» Garage respectively. A «experience microservice» provides tailored API endpoints for each user group. To manage the traffic to an «experience microservice», API proxies provided by MuleSoft are used. In Fig. 2, the «api proxy» A-VehicleUser and an «api proxy» A-Garage for the respective «experience microservice» are shown.

To solve the coarse-grained authorization, the JWT can be inspected by the «api proxy» similar to the PCMFrontend. The corresponding policy is called JWT validation. Before an «api proxy» can introspect the JWT, the PCMFrontend has to include the access token which it previously received from Okta in every HTTP request that is sent to the «api proxy». To include a JWT in an HTTP request, the HTTP authorization header must be set with the JWT set as bearer [FR14; JH12]. If the «api proxy» receives a request without a JWT, the request is immediately returned as unauthorized. Otherwise, the JWT validation policy will perform three checks:

First, it must be validated that the JWT itself has not been tempered with and is not yet expired. A JWT consists of a header, a payload, and a signature [JBS15]. The signature is encrypted by the token issuer (i.e., Okta). To verify the contents of a token, the signature of the token has to be decrypted using a public key. The OAuth 2.0 authorization server can provide the keys to a third party through a JSON Web Key Set (JWKS) endpoint [Jo15]. In this case, the authorization server Default offers a JWKS endpoint which is referenced by the «api proxy» A-VehicleUser and the «api proxy» A-Garage to verify the origin of the JWT. Second, Okta adds the custom claim cid to its JWT which represents the id of the OIDC client that performed the authentication request. An example can be found in line 2 of the JWT displayed in Listing 1. In this case, the value $0oa3ouku57Q89YCEc0x7$ represents the client id of the «oidc client» PCMOC which was used to issue the access token. With this verification, the «api proxy» can verify that the token was requested by the OIDC client of the application PCM. Third, based on the group assigned to the user, the access to an «experience microservice» should either be allowed or denied. Hence, the claim group from the JWT must be evaluated. In the example access token in Listing 1, the user has the groups EveryOne and CaseStudyCompany. Thus, the user should only be allowed to access the «experience microservice» E-VehicleUser. To access the «experience microservice» E-Garage, the group CaseStudyGarage must be present in the JWT.

An exemplary request to receive vehicle information for a specific Vehicle Identification Number (VIN) is presented in Fig. 4. Here, the interaction between involved systems can be seen. The user of the PCMFrontend initiates an HTTP GET request to an endpoint /vin/id

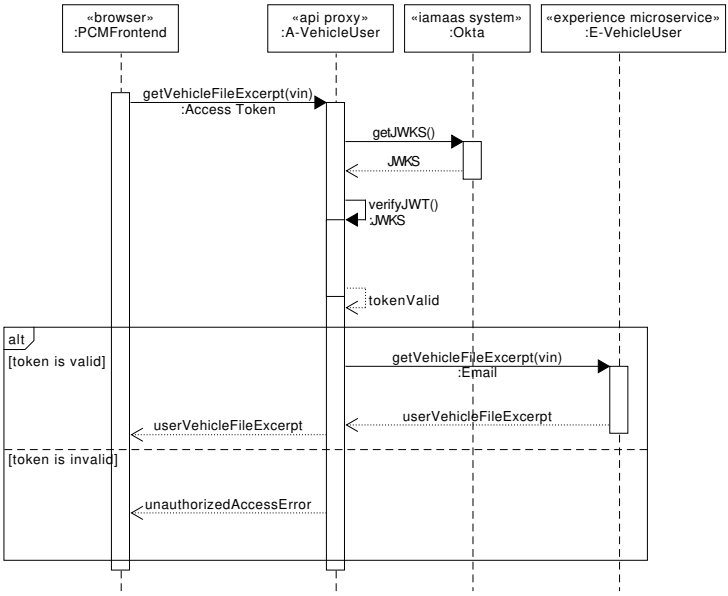


Fig. 4: Excerpt of the Authorization Sequence

which is intercepted by the «api proxy» A-VehicleUser. Before the proxy can apply the JWT validation policy, the JWKS is fetched. Subsequently, the JWT is validated, and the custom claims are checked. If the request is performed by an employee of CSC, the request is forwarded to the «experience microservice» E-VehicleUser and the result is returned to the PCMFrontend. Otherwise, an unauthorized error message is returned to the PCMFrontend.

At this point, the «experience microservice» E-VehicleUser is protected against requests by users outside the CSC and KiC. This leaves the question on how to perform fine-grained authorization, e.g., how to decide which vehicle a user is allowed to access and how to enforce this access. Generally, deciding the access to a resource on a fine-grained level can either be performed inside the application (i.e., business logic) or externalized (i.e., by an external policy enforcement point).

In case of the «experience microservice» E-Vehicle and E-Garage, the authorization decision is performed inside the microservice based on the the email address of the requesting user which is available in the JWT. The experience microservices will perform an internal lookup to check if the email address is allowed to access the VIN. For example, if Alex Twin performs a request for a vehicle file using the access token from Listing 1, the «experience microservice» E-VehicleUser will use the email address alex.twin@csc.com from the token to lookup if the email address is allowed to access the VIN. If so, the request is allowed

and the vehicle file is returned to the «browser» PCMFrontend. Otherwise, an unauthorized error is returned.

5 Conclusion

In this publication, we described the development of a software as a service solution using the cloud products MuleSoft and Okta. The focus is set on securing the application by integrating authentication and authorization mechanisms using the protocols OIDC and OAuth 2.0. The integration of authentication and authorization requires the coordination between systems. The IAMaaS Okta stores the user information and groups and provides the OIDC client which is used by the PCMFrontend. The Angular frontend uses the configured OIDC client to initiate the authentication flow. The PCMFrontend decides which views can be accessed based on the groups of a user. To decide which experience microservices can be accessed by a user, API proxies with a JWT validation policy are used. The fine-grained authorization decisions are performed inside the experience microservices.

Authenticating users in a single page application using OIDC and Angular is a well established topic. However, performing authorization decisions, especially fine-grained authorization decisions, in a microservice architecture must be further researched. In this publication, we performed coarse-grained authorization decisions inside an API proxy by only allowing access to an experience microservice for a certain group. Currently, the decision of who can access a specific vehicle is performed in the logic of the experience microservice. While this is a simple solution, if the requirements for the decision changes, the experience microservice must also be changed due to the tight coupling. Thus, authorization decisions should rather be performed outside the microservice. This also leaves the microservice with its core functionalities. However, externalizing the authorization decision requires additional infrastructure (i.e., decision points) as well as the definitions of access policies. Future research will investigate the structured development, the systematic storage, and the enforcement of authorization policies in applications using the microservice architecture.

References

[AC22] de Almeida, M. G.; Canedo, E. D.: Authentication and Authorization in Microservices Architecture: A Systematic Literature Review. en, Applied Sciences 12/6, p. 3023, Mar. 2022, visited on: 03/31/2022.

[Fo21] Foundation, O.: OWASP Top 10:2021, Publication, OWASP Foundation, 2021.

[FR14] Fielding, R. T.; Reschke, J.: Hypertext Transfer Protocol (HTTP/1.1): Authentication, RFC 7235, June 2014.

[HY17] He, X.; Yang, X.: Authentication and Authorization of End User in Microservice Architecture. en, Journal of Physics: Conference Series 910/, p. 012060, Oct. 2017, visited on: 03/31/2022.

[ID12] Iankoulova, I.; Daneva, M.: Cloud computing security requirements: A systematic review. In: 2012 Sixth International Conference on Research Challenges in Information Science (RCIS). IEEE, pp. 1–7, 2012.

[In22a] Inc., O.: Okta Documentation: Embedded Okta Sign-In Widget fundamentals, tech. rep., [retrieved 05/04/2022], 2022.

[In22b] Inc., O.: The World's #1 Identity Platform | Okta, tech. rep., [retrieved 05/04/2022], Okta Inc., 2022.

[JBS15] Jones, M.; Bradley, J.; Sakimura, N.: JSON Web Token (JWT), RFC 7519, May 2015.

[JH12] Jones, M.; Hardt, D.: The OAuth 2.0 Authorization Framework: Bearer Token Usage, RFC 6750, Oct. 2012.

[Jo15] Jones, M.: JSON Web Key (JWK), RFC 7517, May 2015.

[Mu22a] MuleSoft: API-led connectivity - The next step in the evolution of SOA, https://www.mulesoft.com/lp/whitepaper/api/api-led-connectivity, 2022, visited on: 03/31/2022.

[Mu22b] MuleSoft: MuleSoft | Integration Platform for Connecting SaaS and Enterprise Applications, https://mulesoft.com/, 2022, visited on: 03/31/2022.

[Ne15] Newman, S.: Building Microservices: Designing Fine-grained Systems. O'Reilly Media, Inc., 2015.

[Ne19] Nehme, A.; Jesus, V.; Mahbub, K.; Abdallah, A.: Fine-Grained Access Control for Microservices. In (Zincir-Heywood, N.; Bonfante, G.; Debbabi, M.; Garcia-Alfaro, J., eds.): Foundations and Practice of Security. Vol. 11358, Series Title: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 285–300, 2019, ISBN: 978-3-030-18419-3, visited on: 04/20/2022.

[OM17] OMG: OMG Unified Modeling Language (OMG UML), Superstructure, Version 2.5.1, tech. rep., 2017, visited on: 08/02/2021.

[SBA15] Sakimura, N.; Bradley, J.; Agarwal, N.: Proof Key for Code Exchange by OAuth Public Clients, RFC 7636, Sept. 2015.

[SS75] Saltzer, J.; Schroeder, M.: The protection of information in computer systems. Proceedings of the IEEE 63/9, pp. 1278–1308, 1975.

[Vi11] View, M.; Rydell, J.; Pei, M.; Machani, S.: TOTP: Time-Based One-Time Password Algorithm, RFC 6238, May 2011.

[XJK19] Xu, R.; Jin, W.; Kim, D.: Microservice Security Agent Based On API Gateway in Edge Computing. en, Sensors 19/22, p. 4905, Nov. 2019, visited on: 04/20/2022.

[ZJJ18] Zhao, J. T.; Jing, S. Y.; Jiang, L. Z.: Management of API Gateway Based on Micro-service Architecture. en, Journal of Physics: Conference Series 1087/, p. 032032, Sept. 2018, visited on: 04/20/2022.

D-TOUR : Detour-based point of interest detection in privacy-sensitive trajectories

Maja Schneider[1], Lukas Gehrke[2], Peter Christen[3], Erhard Rahm[4]

Abstract: Data collected through mobile sensors on private and commercial devices can give valuable insights into mobility patterns and facilitate applications such as urban planning or traffic forecasting. At the same time, such data can carry immense privacy risks for the data producers. Stop detection approaches can reveal a person's points of interest (POI) by clustering temporal and spatial features, uncovering private attributes such as home or work addresses. Privacy-preserving mechanisms aim at hiding these POIs, for example via speed smoothing approaches that are able to preserve high data utility. We show experimentally on two real-world data sets that trajectories can contain anomalies that are contained to a certain extent when smoothing a route and are not detected by state of the art stop detection algorithms. We propose a novel attack *D-TOUR* that reveals POIs based on deviations from the optimal route. Our experiments suggest that our proposed attack has similar performance on unprotected data but outperforms the baseline approach, especially when protection is higher and route features become more sparse.

Keywords: Location Privacy; Trajectory data; Inference attack; Points of interest

1 INTRODUCTION

The increasing use of location-based services (LBS) on personal mobile devices or tracking-devices in connected vehicles is generating a wealth of mobility data that is being used by businesses and governments to improve their services and workflows, among other things [BBA15]. In this context, data about movements, such as the trajectories of people and vehicles, are of particular interest. Such data, typically collected using Global Positioning System (GPS) sensors, can provide valuable insight into mobility behavior and enable traffic forecasts and analyses that facilitate, for example, urban planning or cab dispatching [Mo16]. Mobility data drives improved user experience by enhancing mobile apps through personalization or context awareness [LLL16], and offers economic advantages because such data can be monetized and used for targeted advertising.

On the other hand, mobility data are inherently privacy sensitive [Kr07]. For example, private information can be learned by identifying a person's points of interest (POI). Both private POIs, such as home or work addresses, and public POIs, such as stores or museums,

[1] Scads.AI, University of Leipzig, Germany mschneider@informatik.uni-leipzig.de
[2] Scads.AI, University of Leipzig, Germany lg58weky@studserv.uni-leipzig.de
[3] School of Computing, The Australian National University, Canberra, Australia peter.christen@anu.edu.au
[4] Scads.AI, University of Leipzig, Germany rahm@informatik.uni-leipzig.de

doi:10.18420/inf2022_20

are informative for identifying information such as gender, age, education level, or marital status [Zh15b]. Commercially collected movement data, for example in the context of parcel or food deliveries, can also reveal private information about customers, such as the times they are at home. As a recent example, a data leak of a food delivery service could be linked with other data to identify the residence of secret service agents and other politically interesting persons.[5] For corporate customers, POIs can also leak details about business relationships that may be of interest to competitors. Even more can be learned about people's lives if location data can be linked to other sensitive data such as medical records [CRS20].

The privacy risk of trajectory data can be measured, for example by the success probability of attacks on private data, such as tracking or inference attacks, or POI-reconstruction [AS03, GKd11, Kr07, Sh11]. To identify POIs, approaches are generally used that build on temporal and spatial clustering, known as stop detection [HT04]. While this approach relies on the presence of point clusters, a protection against such an attack is to smooth temporal and spatial features [Pr15]. The resulting trajectories hide POIs but preserve to some extent the mobility patterns and thereby ensure data utility, which is a typical challenge for location privacy preserving mechanisms (LPPM) [ABN10].

In this paper we investigate the question of whether such smoothed trajectories still contain enough information for an adversary to uncover private information such as POIs. To this end, we test the attack success of a state of the art stop detection algorithm on smoothed trajectories. We then devise a new attack algorithm *D-TOUR* that builds on analyzing unusual behaviour, such as taking a detour. We argue that trajectory data contains anomalies that can be exploited to identify POIs, and show experimentally on two real-world data sets that these anomalies cannot be sufficiently removed with time- and space-based smoothing approaches. In our experiments, *D-TOUR* is detecting POIs in unprotected trajectories with almost similar performance as the baseline stop detection algorithm. In trajectories protected by the smoothing approach *Promesse* [Pr15], our attack outperforms the baseline, especially when protection is higher and trajectory information becomes more sparse.

2 RELATED WORK

The risk of disclosing a user's POIs and private attributes based on their movement patterns can be measured by the success of an attack on the trajectory, such as inference attacks that infer POI locations, also called stop detection [HT04]. To detect POI locations, most mechanisms work with temporal and spatial features. In a first step, potentially interesting locations are identified as so-called stay-points or stop-points which are then aggregated into stay-regions to infer a POI location. Ashbrook and Starner [AS03] define stay-points as those trajectory points that have a minimum temporal distance from their predecessor, assuming that the large time gap stems from lost GPS-signal in buildings. Stay-points are clustered with

[5] Philipp Bovermann, "Lieferdienst-Datenleck enttarnt russische Geheimdienstmitarbeiter". Online at: https://www.sueddeutsche.de/wirtschaft/datenleak-russland-lieferdienst-bellingcat-1.5561080, 2022

a variant of k-means. However, this approach suffers from false positives due to GPS signal loss in narrow urban areas and fails to identify outdoor locations. Patterson et al. [Pa03] and Liao et al. [Li07] take into account derived features and background knowledge about bus schedules, to learn transportation mode changes that indicate a POI location. This approach relies on such changes and might miss POIs, such as indoor stop locations.

Simple heuristic approaches were mentioned by Gambs et al. [GKd11] to detect POIs, for example choosing the last stop of a trajectory before midnight to identify the home address of a user. The authors reason that POIs can be detected not only by observing spatially correlated events, but also the absence of data in a certain time period which can hint to a POI or private activity. They detected POIs of taxi drivers, for example, by examining when the GPS sensor was not active. Similarly, it is possible to identify muslim cab drivers by comparing their pause timing with praying times.[6]

Several clustering approaches exist for POI location inference that are based on analyzing temporal and spatial features [GKd11, HT04, Ka05, Zh04]. Primault et al. [Pr14] use for their analyses a variation of the Density-Joinable cluster (DJ-Cluster) algorithm [Zh04]. Their adapted algorithm works by first detecting so-called stays, which are areas of a certain diameter in which a person stayed for a certain time period [HT04]. For this the algorithm considers a distance threshold that defines the maximal diameter of a stay area and a time threshold defining the minimum time duration that has to be spent in every stay. In a second step, the stays are grouped into clusters, taking into account a minimum number of stays in a cluster and a merge threshold that defines the maximum distance below which two different clusters are merged into one.

A variety of defenses have been developed against inference attacks. Gambs et al. [GKd11] develop and test the resilience of a clustering mechanism against sanitization measures, such as sampling, a protection algorithm which fuses multiple routes of a user into a single route, and perturbation, which modifies route points by adding spatial Gaussian noise. The authors find perturbation to have the highest defense capabilities, but at the same time a substantial decrease in utility. In general, an LPPM aims at keeping the utility of the data high for the purpose of further analyses.

As clustering algorithms rely on temporal or spatial clusters to work with, an efficient defense against inference attacks based on clustering is to remove points from these clusters, for example by using smoothing approaches. An example for this approach is *Promesse* [Pr15], which keeps distortion in protected trajectories low and utility high. However, the question remains, if it is still possible to infer POI locations based on distinctive characteristics of a trajectory, such as movement patterns, sparsity, or unexpected movement behaviour. In our study we investigate this question and present a novel inference attack that exploits unexpected behaviour in movements to detect POI locations.

[6] L. Franceschi-Bicchierai, “Redditor cracks anonymous data trove to pinpoint muslim cab drivers”. Online at: http://mashable.com/2015/01/28/redditor-muslim-cab-drivers, 2015.

3 FORMALIZATION

To model the investigated problem we rely on the formalization introduced by Shokri et al. [Sh11]. The authors describe the connection between a *mobile user* who produces privacy-sensitive trajectory data while moving, an *adversary* with certain capabilities and knowledge who aims at exploiting the observed data, and an *LPPM* to protect this data from the adversary. We now describe these three concepts.

Mobile user: We consider a user $m \in \mathcal{M}$ moving in a *region* $\mathcal{R}$ during an observed *time period* $\mathcal{T}$. The region consists of distinct locations (or points) $r \in \mathcal{R}$. The time period has a certain resolution, such as minutes, and is defined as a list of timestamps $[t^{start}, t^{start+1}, \ldots, t^{end}]$ that mark, for example, the start of such an interval. An *event* $e \in \mathcal{E}$ captures a single measurement of the user's location r at time $t \in \mathcal{T}$, and is described by a tuple $e = \langle m, r, t \rangle$. Successive events generated by one user form a *trajectory* (or route) $L = [e_1, e_2, \ldots, e_n]$. Such a trajectory can be generated, for example, when a moving person is tracked every few minutes on their cell phone while using a navigation app.

Adversary: We focus on offline privacy, which means that an adversary only looks at the full data set of trajectories. We do not consider online privacy, which investigates the privacy risk that exists at the time of data collection and transmission. The trajectories observed by the adversary may have been previously manipulated by a privacy preserving mechanism.

We assume the adversary knows the mobile user's transport mode (such as walking, car, subway) or is able to deduce it. Indeed, this information may not be coherent in one trajectory, for example when a user travels partly by foot and subway. The adversary also has knowledge of the optimal (e. g. fastest) route between arbitrary locations in region R, taking into account the known mode of transportation. To achieve this, the adversary queries navigation services that calculate the fastest road connection between two locations.

Location Privacy Preserving Mechanisms (LPPM): We consider *Promesse* [Pr15] as the privacy preserving mechanism, which smoothes temporal and spatial features in a trajectory in order to avoid the formation of point clusters of temporally and geographically close events. This algorithm was shown to provide a good trade-off between protection against stop detection attacks and preservation of data utility.

First, the algorithm regularly samples locations along the shape that is formed when successive points of the original route are connected. Successive sampled points have a distance of at least α, the so-called smoothing distance. Then, the first and last point of the route are removed to avoid leaking potentially private information, because trajectory endpoints are likely meaningful [GKd11]. Finally, the timestamps of the sampled points are set to be equally distributed between the original start and end timestamps.

4 DETOUR-BASED POI DETECTION

State of the art approaches in stop detection rely on the presence of stay-points, which are multiple points that are geographically or temporally clustered [HT04]. These clusters result from extended stays by a user in a particular area, which in turn may indicate a potentially privacy-sensitive POI. An intuitive approach to protecting POIs is to remove points from these clusters in order to hide stay-points and thus prevent inference attacks. At the same time, this should be done in a way that does not interfere too much with mobility patterns, so that the benefits are maintained and the data can still be used for mobility analyses.

However, an adversary can use additional knowledge about typical user movements to reveal anomalies in the trajectories that become visible when comparing privacy-sensitive trajectories and known patterns. Following this approach we propose a novel attack to identify POIs in trajectories, named *D-TOUR*, that investigates whether a trajectory contains unnecessary detours. This idea is based on the reasoning that in order to reach a POI, it is usually necessary to leave the path that would be ideal to attain the final destination of the route. The attack is outlined in Algorithm 1 and consists of four steps, described below. Figure 1 illustrates the functionality of the attack on an unprotected and a protected route.

(1) Map-matching The adversary considers a single trajectory and first maps the locations of each of its events to the road network (lines 3 to 5). This so-called map-matching is a typical pre-processing routine of raw trajectories, because data collected with GPS can be inaccurate and off roads [Zh15a]. This step is optional and not shown in Figure 1.

(2) Subset selection In a second step (lines 7 to 14), the attacker successively selects several events from the trajectory, starting with the first event and maintaining a certain selection distance δ between successive selected events. The parameter δ needs to be adjusted depending on the tracking interval (the time passed between consecutive events) of the data. If δ is too short, detours will not be detected, whereas if it is too long, parts of a route will be wrongly interpreted as a detour. The adversary adds the last route point to the selection in case it was not selected. In Figure 1 the selected points are highlighted by black circles.

(3) Distance to optimal route calculation In the third step (lines 16 to 26), the adversary aims to find deviations between the user's route and an optimal route. For this, the adversary first connects successive points from the selection (line 18) using the function $OPT(e_i, e_j)$ that returns the way points of the optimal route between the locations of the events e_i and e_j. For example, the optimal route can be interpreted as the fastest road connection and queried by public routing services[7], taking into account the user's known means of transport. In Figure 1 the fastest route is shown as a gray line.

Then, for each point of the user's trajectory, the adversary aims to calculate the shortest distance to the optimal route. A routing service's response usually contains way points that

[7] In our experiments, calculation of the fastest route is done with openrouteservice.org.

Algorithm 1 D-TOUR: Detour-based POI detection

Require: route $L = [e_1, e_2, \ldots, e_n]$, selection distance $\delta \geq 0$, sampling distance $\sigma \geq 0$, acceptable distance $\gamma \geq 0$

function D-TOUR($L, \delta, \sigma, \gamma$)

// Step 1

  for $i = 1, 2, \ldots, n$ **do** ▷ Map each route point to road network.

    $e_i \leftarrow MAP(e_i)$

  end for

// Step 2

  $c \leftarrow 1$ ▷ Take a selection C of point indices from the route,

  $C \leftarrow [1]$ starting by first, keeping a minimal distance δ

  for $i = 2, 3, \ldots, n$ **do** between consecutive events.

    if $DIST(e_c, e_i) \geq \delta \vee i = n$ **then**

      $c \leftarrow i$

      Add i to C

    end if

  end for

// Step 3

  $X \leftarrow []$ ▷ The exceedances of the acceptable distance.

  for $i = 2, 3, \ldots, |C|$ **do** ▷ Loop over selection.

    $O \leftarrow OPT(e_{C[i-1]}, e_{C[i]})$ ▷ Calculate optimal route O between consecutive events.

    $S \leftarrow SAMPLE(O, \sigma)$ ▷ Take regular samples S with distance σ from O.

    $L^c = [e_{C[i-1]+1}, e_{C[i-1]+2}, \ldots, e_{C[i]}]$ ▷ The route section between selected points.

    for e in L^c **do** ▷ Loop over route section.

      $d = \min_{s \in S} DIST(e, s)$ ▷ Calculate distance d to nearest sample.

      $x \leftarrow \max(d - \gamma, 0)$ ▷ Calculate by how much d exceeds

      Add x to X the acceptable distance γ.

    end for

  end for

// Step 4

  $I \leftarrow []$ ▷ The point indices of a POI candidate region.

  $P \leftarrow []$ ▷ The identified POIs.

  $b \leftarrow X[1] > 0$ ▷ b is True, if a POI candidate region is detected.

  for $i = 1, 2, \ldots, n$ **do** ▷ Loop over route.

    if $X[i] > 0$ **then** ▷ If a route point exceeds the acceptable distance,

      Add i to I consider it a POI candidate.

      $b \leftarrow True$

    end if

    if $b \wedge (X[i] = 0 \vee i = n)$ **then** ▷ If a closed region of connected POI candidates

      $p = \arg\max_{j \in I} X[j]$ is detected, the candidate p with the greatest

      Add p to P distance exceeding is considered a POI.

      $b \leftarrow False$

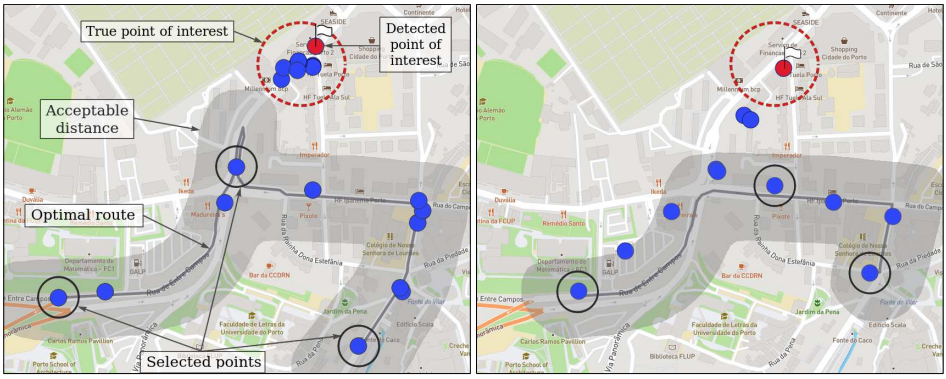
      $I \leftarrow []$

    end if

  end for

  return P ▷ Return the identified POIs.

end function



(a) Unprotected route (b) Protected route using *Promesse* (α = 200 meters)

Fig. 1: (a) An example of a *D-TOUR* attack on a privacy-sensitive route. Based on a selection of points from the route, *D-TOUR* calculates the optimal route and considers points that deviate from it by more than a certain acceptable distance as POI candidates. Contiguous regions of POI candidates are interpreted as a detour. The furthest point in such a region is identified as a POI. (b) The protection with *Promesse* does not conceal the detour and *D-TOUR* can still identify a point within a certain circular area around the true POI location.

can be far apart. To find a point's closest neighbour on the optimal route it is necessary to sample regularly spaced points from the route in short intervals. For this the function $SAMPLE(L, \sigma)$ moves along the route L on a straight line between consecutive points (line 19) and takes samples from this path in regular intervals of length σ by measuring the location. Each sample forms an event without time and is added to a list. Small values of σ favor a more accurate distance calculation but have greater computational demand.

Using the function $DIST(e_i, e_j)$, the adversary then calculates for each route point e_i the Euclidean distance to each of the sampled points e_j from the optimal route and selects the shortest distance in each case (line 22). At the same time the adversary sets a threshold γ as an acceptable distance between a point and the optimal route and keeps track of by how much each of the calculated distances exceeds this value (lines 23 and 24). The acceptable distance γ is outlined as a gray area around the optimal route in Figure 1.

(4) POI detection In the final step, the adversary considers each route point that exceeds the acceptable distance γ as a POI candidate (lines 32 to 35). Consecutive POI candidates then form a candidate region and indicate a detour. In Figure 1 these candidates are those points, that lie outside of the gray area. For each of these candidate regions, the point with the greatest distance to the optimal route is considered a POI (lines 36 to 41). Figure 1 shows a true POI as a circular area around its actual location. A POI detected by an attack is marked with a flag and considered a true positive if it lies within the circular area.

Tab. 1: Data sets used in experimental evaluation.

Data set	Location	Tracking interval	Timespan	No. routes
Cabspotting	San Francisco (USA)	60 sec	17 May - 10 Jun 2008	12,000
Porto Taxi	Porto (Portugal)	15 sec	1 July - 2 Aug 2013	25,800

5 Experiments

In this section we describe an evaluation of our proposed *D-TOUR* method on two real-world data sets, and discuss our findings.

Data sets: We run experiments on two data sets as shown in Table 1. *Cabspotting* [PSG09] is a collection of taxi routes and contains the occupancy of a cab which we use to derive the pick-up and drop-off points of taxi customers, which serve as the POI ground truth. The *Porto Taxi* [Mo13] data set only includes trips made by occupied cabs and does therefore not contain continuous routes. Trip endpoints correspond to the customer pick-up and drop-off locations and are regarded as POIs. We clean the Porto Taxi data set from trips that take less than a minute or that contain presumably incorrect measurements that indicate speeds of more than 100 kilometers per hour.

We then match trips with the same taxi identifier and day and join them to form a continuous route, either by a direct link, the fastest route, or a random walk. To ensure a realistic matching we compare, among other things, the gap duration between trip endpoints to the duration of their shortest road connection. If the information is not coherent, we create a random walk that satisfies the duration constraints on the endpoints. A random walk is searched for a maximum of 500 attempts. If no valid route is found, a new route is created from the trip. A new route is also created if the gap is longer than 60 minutes.

Experimental setup: We first evaluate our proposed *D-TOUR* algorithm on the raw non-private trajectories of each data set. In a second step we then protect the data with *Promesse* [Pr15] to smooth temporal and spatial features of the routes and hide POIs. We argue that a trajectory protected in such a way still contains anomalies that indicate POIs. Therefore we evaluate POI detection on the protected routes with our detour-based detection approach[8]. We run experiments with data protected by smoothing distances α of 200, 300 and 400 meters.

In both steps we compare the performance of *D-TOUR* to a variation of the DJ-Cluster algorithm as described by Primault et al. [Pr14], which is a commonly used stop detection baseline [HT04, Li08]. In this algorithm, first, stays are extracted which are then aggregated into POIs. The algorithm takes as parameters a time threshold Δt which corresponds to the minimum time that has to be spent in a stay, and a distance threshold Δl which represents the maximum diameter of the stay area.

[8] For implementation see: https://github.com/majaschneider/detour-detection.

Evaluation metrics: We measure the performance of a stop detection algorithm based on how many of a route's POIs are correctly detected and whether detected POIs indeed denote a true stop. Following the approach of Primault et al. [Pr14] we define a detected stop and a true stop to be equal when their geographical distance is smaller than a parameter β. With this definition we can count the number of correctly detected true stops (true positives), as well as the number of stops that have not been detected (false negatives), and the number of stops that have been detected but are in fact not a true stop (false positives). In our experiments we consider $\beta = 200$ meters.

We calculate *Recall*, *Precision* and *F-Score* to evaluate the performance of stop detection. Recall denotes the ratio of correctly classified stops to all true stops. If it is low, a high number of stops have not been detected. Precision, on the other hand, is the ratio of correctly detected stops to the total number of stops that have (correctly or incorrectly) been identified as true stops. A low precision value indicates therefore a high false positive rate. Both metrics are combined in the F-Score, which is defined as their harmonic mean.

Experimental results: As opposed to Primault et al. [Pr15] we do not assume fixed parameters for stop detection but tune them on a subset of 2000 routes from each data set to obtain the best settings for both DJ-Cluster and *D-TOUR*. For DJ-Cluster we test time thresholds Δt of 1, 3, 5, 10 and 15 minutes, and distance thresholds Δl of 200, 300, 400, 500 and 1000 meters. The best parameters showed to be $\Delta t = 1$ minute and $\Delta l = 200$ meters for the Cabspotting data set which is amongst the four best settings for the Porto Taxi data set as well (only 3% lower F-Score). Therefore we choose these settings for both data sets.

The success of *D-TOUR* depends heavily on how well the spatial distance σ captures the outline of the route and avoids selected points to be located close to a POI. Therefore we also test combined results of multiple spatial distances and consider only those POIs that have been correctly identified in most settings. In our experiments we tested all such combinations of spatial distances σ of 100, 110, 310, 320, 410, 420, 610, 620 and 1000 meters together with acceptable distances γ of 20, 30, 50, 70, 100 and 150 meters, both with and without map-matching. However, the best parameters (for both data sets) were a single spatial distance $\sigma = 620$ meters and an acceptable distance $\gamma = 20$ meters using no map-matching.

The results of our experiments are presented in Table 2. Both attacks are capable of capturing POIs in unprotected data. The overall performance is low with a maximum F-Score of 0.45, indicating that POI detection is a difficult task. Both attacks have a comparable F-Score in the Cabspotting dataset, while DJ-Cluster in the Porto-Taxi dataset performs 6% better.

As expected, with increasing protection (i. e. larger smoothing parameter α) both attacks are less successful with regards to F-Score. The performance of DJ-Cluster drops for the lowest protection of $\alpha = 200$ meters by up to 22%. *D-TOUR* also shows a performance drop but is able to achieve an F-Score which is, compared to the baseline, 13% better for Cabspotting and 3% better for Porto Taxi data. While in general all performance measures decrease with higher privacy protection, an exception is recall for *D-TOUR* which is improving

Tab. 2: Experimental results.

Algorithm	Protection (*Promesse*)	Cabspotting			Porto Taxi		
		Recall	Precision	F-Score	Recall	Precision	F-Score
DJ-Cluster	Unprotected	0.43	0.47	0.45	0.53	0.34	0.41
	$\alpha = 200m$	0.20	0.34	0.25	0.42	0.13	0.19
	$\alpha = 300m$	0	0	0	0	0	0
	$\alpha = 400m$	0	0	0	0	0	0
D-TOUR	Unprotected	0.51	0.39	0.44	0.38	0.32	0.35
	$\alpha = 200m$	0.67	0.27	0.38	0.43	0.15	0.22
	$\alpha = 300m$	0.52	0.27	0.35	0.15	0.10	0.12
	$\alpha = 400m$	0	0	0	0	0	0

significantly for the lowest protection, rising by 15% (Cabspotting) and 5% (Porto Taxi). The *D-TOUR* algorithm correctly detects more POIs but also makes more mistakes, leading to an overall reduced F-Score. With $\alpha = 300$ meters DJ-Cluster is not able to detect any POIs anymore, whereas *D-TOUR* still achieves an F-Score of up to 0.35. Only with $\alpha = 400$ meters *D-TOUR* is also not able to capture any POIs anymore.

Discussion: POIs found with *D-TOUR* could be further evaluated using additional knowledge. For example, an adversary can try to label POIs using public map data and estimate the probability of correctness using contextual knowledge (such as the source of the data). Also, different types of attacks could be combined to increase the confidence in the results.

However, a privacy preserving mechanism must guarantee that POIs are sufficiently protected, which is why even a low recall is a potential threat to privacy. Further addition of noise (e. g. higher smoothing value α) can help mitigate this issue but is ultimately limited as it deteriorates utility. Instead, a more targeted elimination of anomalies can be beneficial to ensure privacy is protected and utility stays high.

In this context, *Promesse* introduces a spatial error. This can not only lead to a decrease of utility but can be a reason for the performance drop of *D-TOUR* when protected data is attacked. The attempted mitigation with map-matching did not provide an improvement in our experiments, which indicates that the spatial error introduced by *Promesse* is too large.

Visual inspection of the data shows that the street layouts in the investigated cities differ, which can also have an impact on the spatial error in protected data. Even though the tracking interval of San Francisco cabs is larger, the grid layout of the streets in this city keeps the spatial error low. We hypothesize that this is the reason for the better performance of *D-TOUR* in the Cabspotting data set compared to the Porto Taxi data set.

In addition, the fact that start and endpoints of Porto Taxi routes are considered POIs as well contributes to the worse performance of *D-TOUR* on this data set. Heuristics could be

used as additional treatment, such as the begin-end-heuristic [GKd11], which simply picks the first and last trajectory point as POIs. This is a low-cost approach based on the idea that start and endpoints of routes are usually meaningful.

According to the authors of *Promesse* [Pr15], the smoothing parameter α is key to hiding POIs that would be otherwise extracted with a parameter $\Delta l \leq \alpha$ in a DJ-Cluster attack. Our experimental results are consistent with this reasoning as POI detection performance of DJ-Cluster drops significantly with $\alpha = 200$ meters and $\Delta l = 200$ meters, and even further with $\alpha = 300$ meters. *D-TOUR*, on the other hand, is still able to detect POIs in this setting, which shows that it is not sufficient to evaluate privacy requirements based only on DJ-Cluster.

6 CONCLUSION AND FUTURE WORK

We showed experimentally on two real-world data sets that POIs in trajectories of mobile users can be regarded as an anomaly from typical movement behaviour, which can be detected by searching for detours in a route. We proposed a novel attack, *D-TOUR*, that compares a sensitive trajectory to the fastest route between a selection of its points and chooses the furthest point of such a deviation from the optimal route as a POI. Compared to DJ-Cluster, a state of the art stop detection algorithm, our attack achieves similar performance on unprotected data but outperforms the baseline when privacy protection was increased.

As future work we will focus on investigating the influence of the actual transport mode (such as walking, car, or subway) on the attack success. We also plan to investigate more diverse data sets containing different data producers, such as logistics companies or tourists that show different moving behaviour and approach other types of POIs. We furthermore plan to investigate whether results can be improved with a revised map-matching approach, that matches points preferentially with larger roads to reduce the spatial error.

Acknowledgements This work is partially funded by the German Federal Ministry of Education and Research under grant BMBF 01IS18026B at ScaDS.AI Dresden/Leipzig.

Bibliography

[ABN10] Abul, Osman; Bonchi, Francesco; Nanni, Mirco: Anonymization of moving objects databases by clustering and perturbation. Inf. Syst., 35(8):884–910, 2010.

[AS03] Ashbrook, Daniel; Starner, Thad: Using GPS to learn significant locations and predict movement across multiple users. Pers. Ubiquitous Comput., 7(5):275–286, 2003.

[BBA15] Ben Ayed, Abdelkarim; Ben Halima, Mohamed; Alimi, Adel M.: Big data analytics for logistics and transportation. IEEE ICALT, pp. 311–316, 2015.

[CRS20] Christen, Peter; Ranbaduge, Thilina; Schnell, Rainer: Linking Sensitive Data. Springer, Heidelberg, 2020.

[GKd11] Gambs, Sèbastien; Killijian, Marc Olivier; del Prado Cortez, Miguel Núñez: Show me how you move and I will tell you who you are. Trans. Data Priv., 4(2):103–126, 2011.

[HT04] Hariharan, Ramaswamy; Toyama, Kentaro: Project lachesis: Parsing and modeling location histories. In: Proc. GISci. pp. 106–124, 2004.

[Ka05] Kang, Jong Hee; Welbourne, William; Stewart, Benjamin; Borriello, Gaetano: Extracting places from traces of locations. ACM SIGMOBILE Mob. Comput. Commun. Rev., 9(3):58–68, 2005.

[Kr07] Krumm, John: Inference attacks on location tracks. In: Proc. 5th Int. Conf. Pervasive Comput. pp. 127–143, 2007.

[Li07] Liao, Lin; Patterson, Donald J.; Fox, Dieter; Kautz, Henry: Learning and inferring transportation routines. Artificial Intelligence, 171(5-6):311–331, 2007.

[Li08] Li, Quannan; Zheng, Yu; Xie, Xing; Chen, Yukun; Liu, Wenyu; Ma, Wei Ying: Mining user similarity based on location history. In: Proc. ACM SIGSPATIAL GIS. pp. 298–307, 2008.

[LLL16] Liu, Xin; Liu, Yong; Li, Xiaoli: Exploring the Context of Locations for Personalized Location Recommendations. In: IJCAI. pp. 1188–1194, 2016.

[Mo13] Moreira-Matias, Luis; Gama, Joao; Ferreira, Michel; Mendes-Moreira, Joao; Damas, Luis: Predicting taxi-passenger demand using streaming data. IEEE Trans. Intell. Transp. Syst., 14(3):1393–1402, 2013.

[Mo16] Moreira-Matias, Luís; Gama, João; Ferreira, Michel; Mendes-Moreira, João; Damas, Luis: Time-evolving O-D matrix estimation using high-speed GPS data streams. Expert Syst. Appl., 44:275–288, 2016.

[Pa03] Patterson, Donald J; Liao, Lin; Fox, Dieter; Kautz, Henry: Inferring High-Level Behavior from Low-Level Sensors. In: Proc. UbiComp. pp. 73–89, 2003.

[Pr14] Primault, Vincent; Mokhtar, Sonia Ben; Lauradoux, Cedric; Brunie, Lionel: Differentially Private Location Privacy in Practice. In: Proc. MoST. 2014.

[Pr15] Primault, Vincent; Ben Mokhtar, Sonia; Lauradoux, Cédric; Brunie, Lionel: Time distortion anonymization for the publication of mobility data with high utility. IEEE Trustcom/BigDataSE/ISPA, 1:539–546, 2015.

[PSG09] Piórkowski, Michal; Sarafijanovic-Djukic, Natasa; Grossglauser, Matthias: A parsimonious model of mobile partitioned networks with clustering. In: Proc. of COMSNETS. 2009.

[Sh11] Shokri, Reza; Theodorakopoulos, George; Le Boudec, Jean Yves; Hubaux, Jean Pierre: Quantifying location privacy. In: Proc. IEEE Symp. Security Privacy. pp. 247–262, 2011.

[Zh04] Zhou, Changqing; Frankowski, Dan; Ludford, Pamela; Shekhar, Shashi; Terveen, Loren: Discovering personal gazetteers: An interactive clustering approach. In: Proc. ACM SIGSPATIAL GIS. pp. 266–273, 2004.

[Zh15a] Zheng, Yu: Trajectory data mining: An overview. ACM Trans. Intell. Syst. Technol., 6(3):1–41, 2015.

[Zh15b] Zhong, Yuan; Yuan, Nicholas Jing; Zhong, Wen; Zhang, Fuzheng; Xie, Xing: You are where you go: Inferring demographic attributes from location check-ins. In: Proc. WSDM. pp. 295–304, 2015.

Messung der Datenminimierung für den Beschäftigtendatenschutz am Beispiel von Standortdaten

Verifikation der datenschutzrechtlichen Anforderungen an die Datenminimierung mithilfe von Metriken[1]

Janine Schleper[2], Matthias Kohn[3], Paulina Jo Pesch[4], Ulrich Waldmann[5], Thomas Kunz[6]

Abstract: Metriken zur automatisierten Verifizierung der Umsetzung des Datenminimierungsgebots können Unternehmen darin unterstützen, ihren Kontrollpflichten nachzukommen und das Vertrauen ihrer Beschäftigten in eine datenschutzkonforme Verarbeitung ihrer Daten zu stärken. Um sich durch eine Metrik einer komplexen datenschutzrechtlichen Anforderung wie dem Gebot der Datenminimierung anzunähern, sind geeignete Datenquellen sinnvoll miteinander zu kombinieren. Diese Arbeit skizziert grundlegende Metriken zur Kontrolle der Datenminimierung anhand eines Anwendungsszenarios im Bereich der agilen Personaleinsatzplanung und beschreibt die prototypische Umsetzung dieser Metriken.

Keywords: Automatisierte Kontrolle; Datenschutz; Messquellen; Metrik; Datenminimierung

1 Herausforderungen und Ziel

Metriken bieten die Möglichkeit, den Grad der Umsetzung von Datenschutzvorschriften wie dem Datenminimierungsgebot auf Grundlage von Messdaten automatisiert zu verifizieren. Sie können Unternehmen dabei unterstützen, von mehreren Optionen die datenschutzfreundlichste zu ermitteln oder über die Einhaltung von Datenschutzvorschriften Rechenschaft abzulegen – und dies sowohl gegenüber Aufsichtsbehörden und betrieblichen

[1] Das diesem Beitrag zugrunde liegende Vorhaben EduMiDa wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 16KIS1361K, 16KIS1362 und 16KIS1363 gefördert. Die Verantwortung für den Inhalt liegt bei den Autor*innen. Michael Koddebusch gebührt unser Dank für hilfreiche Hinweise.

[2] Universität Bremen, Institut für Informations-, Gesundheits- und Medizinrecht (IGMR), Universitätsallee, 28359 Bremen, Deutschland, schleper@uni-bremen.de

[3] Universität Bremen, Institut für Informations-, Gesundheits- und Medizinrecht (IGMR), Universitätsallee, 28359 Bremen, Deutschland, kohn@uni-bremen.de

[4] WWU Münster, Institut für Wirtschaftsinformatik, Leonardo-Campus 3, 48149 Münster, Deutschland, paulina.pesch@uni-muenster.de

[5] Fraunhofer SIT, Cloud Computing, Identity Privacy (CIP), Rheinstr. 75, 64295 Darmstadt, Deutschland, ulrich.waldmann@sit.fraunhofer.de

[6] Fraunhofer SIT, Cloud Computing, Identity Privacy (CIP), Rheinstr. 75, 64295 Darmstadt, Deutschland, thomas.kunz@sit.fraunhofer.de

doi:10.18420/inf2022_21

Datenschutzbeauftragten, als auch gegenüber Betroffenen. Gerade für die Verbesserung und Kontrolle des Beschäftigtendatenschutzes durch Unternehmen bieten sich Metriken an [3]. Dies gilt insb. für das in Art. 5 Abs. 1 lit. c DSGVO verankerte Datenminimierungsgebot (im Einzelnen s. Abschnitt 2). Unternehmen, die entgegen der Vorschrift zu viele Daten verarbeiten, drohen gemäß Art. 83 Abs. 5 lit. a DSGVO hohe Bußgelder.

2 Anwendungsszenario und Anforderungen

In der Logistikbranche verspricht die Nutzung von Echtzeit-Standortdaten wirtschaftliche Vorteile. So kann zur Aufrechterhaltung des Betriebsablaufs eine agile Personaleinsatzplanung (PEP) erforderlich sein und die Nutzung von Standortdaten Kosten senken. Wenn auf einem großen Betriebsgelände eine technische Anlage ausfällt, müssen nicht nur Beschäftigte ermittelt werden, die qualifiziert sind die Anlage zu reparieren und sich im Dienst befinden, sondern ist auch die Identifikation der Beschäftigten sinnvoll, die die Anlage möglichst schnell erreichen können. So können Beschäftigte eingesetzt werden, für deren Einsatzortverlegung möglichst geringe Kosten anfallen. Die Verarbeitung von Standorten muss dem Datenminierungsgebot nach Art. 5 Abs. 1 lit. c DSGVO genügen. Danach müssen personenbezogene Daten dem Zweck *angemessen* und *erheblich* sowie auf das für die Zwecke der Verarbeitung *notwendige Maß* beschränkt sein. Dies erfordert eine möglichst weitgehende Begrenzung der Verarbeitung personenbezogener Daten sowohl im Hinblick auf Quantität, als auch auf Qualität.[7]

Die Datenverarbeitung ist dem Zweck *angemessen*, wenn ein objektiver sachlicher Bezug hinsichtlich Funktion, Inhalt und Umfang zum damit verfolgten Zweck vorliegt.[8] Die Erhebung der Standortdaten muss also zum Zeitpunkt der Datenabfrage in konkretem Bezug zur Beschäftigteneinsatzplanung stehen. Das wäre nicht der Fall, wenn der Beschäftigte für die geplante Tätigkeit erkennbar nicht geeignet ist, etwa eine Reinigungskraft für die Reparatur einer Industrieanlage.

Erheblichkeit für den Zweck setzt voraus, dass die erhobenen Daten objektiv dazu geeignet sind, den mit der Datenverarbeitung verfolgten Zweck zu erreichen.[9] Dafür müssen die erhobenen Daten zumindest für einen Teilaspekt des verfolgten Zwecks bedeutend sein und dessen Erreichung dienen.[10] Die Erhebung des Standorts einer Person, die für die vorgesehene Aufgabe zwar qualifiziert ist, der aber gerade keine dienstliche Anweisung erteilt werden kann, ist zwar angemessen, jedoch nicht erheblich.

Beschränkt auf das *notwendige Maß* ist die Datenverarbeitung dann nicht mehr, wenn der verfolgte Zweck auch ohne die Datenverarbeitung erreicht werden kann.[11] Es sind nur

[7] *Frenzel* in: [6], Art. 5 DSGVO, Rn. 34.
[8] *Herbst* in: [5], Art. 5 DSGVO, Rn. 57; *Roßnagel* in: [7], Art. 5 Rn. 119.
[9] *Schantz* in: [8], Art. 5 DS-GVO, Rn. 24; *Frenzel* in: [6], Art. 5 DSGVO, Rn. 35.
[10] *Roßnagel* in: [7], Art. 5 Rn. 120.
[11] *Roßnagel* in: [7], Art. 5 Rn. 121.

solche Daten zu erheben, die zur Bestimmung des aktuellen Standorts der qualifizierten, diensthabenden Beschäftigten auf dem Werksgelände notwendig sind, um die Person ausfindig zu machen, die die Anweisung zur Reparatur der Anlage erhalten soll. Dies erfordert nicht die Nutzung von Standortdaten zur Ermittlung weiterer Informationen wie etwa zur Arbeitsleistung, Körperhaltung, Bewegung, zum Verhalten in den Pausen oder dem Gesundheitszustand der Beschäftigten, sodass eine Erhebung dieser Daten nicht notwendig und damit nicht mit dem Datenminimierungsgebot vereinbar wäre.

3 Metrikenentwicklung und Messpunkte

Überprüfen lässt sich der Grad der Einhaltung des Datenminimierungsgebots durch Metriken, d. h. präzise definierte Methoden zur Messung bestimmter Attribute eines (Teil-)Systems [2]. **Datenschutzmetriken** dienen speziell der Messung des Grades der Umsetzung datenschutzrechtlicher Ziele wie z. B. Datensicherheit oder Datensparsamkeit. Unternehmen können sich solcher Metriken bedienen, um bei der Wahl zwischen mehreren Lösungen oder der Konfiguration der von ihnen eingesetzten Software die datenschutzfreundlichste Option zu ermitteln oder um Aufsichtsbehörden und Betroffenen gegenüber ein hohes Datenschutzniveau zu demonstrieren.

Einerseits lassen sich Metriken *„bottom up“* ausgehend von den tatsächlich verfügbaren Messdaten entwickeln [1]. Bei Datenschutzmetriken, mittels derer Unternehmen eigene Systeme überprüfen und die eigene Compliance nachweisen, können – anders als bei Metriken zur Überprüfung der Compliance Dritter wie z. B. Cloud-Anbieter [4] – Daten direkt im System erhoben und Messverfahren in dieses implementiert werden. Andererseits lassen sich Metriken *„top down“* ausgehend von dem Ziel entwickeln, für das der Grad der Umsetzung durch die Metrik ermittelt werden soll [1]. Für Datenschutzmetriken besteht dieses in der jeweilig überprüften datenschutzrechtlichen Anforderung, z. B. Datensicherheit oder Datenminimierung. Dabei sind in einem ersten Schritt aus der unspezifischen, abstrakten datenschutzrechtlichen Anforderung konkrete Maßnahmen abzuleiten [4] und dann Metriken zur Messung des Grades der Umsetzung der Maßnahmen zu bilden.

Metriken lassen sich *„top down“* ausgehend von den in Abschnitt 2 beschriebenen einzelnen **Anforderungen** des Datenminimierungsgebots ableiten: Die Daten müssen einen objektiv-sachlichen Bezug zum Verarbeitungszweck aufweisen (Angemessenheit) und ihre Verarbeitung muss zur Zweckerreichung geeignet sein (Erheblichkeit) und auf das notwendige Maß beschränkt sein.

Hieraus lassen sich Maßnahmen entwickeln, deren Einhaltung messbar ist. Das Datenminimierungsgebot misst Anforderungen an die Datenverarbeitung an ihrem Zweck bzw. Anlass. Deshalb lässt sich bloß auf Grundlage der Häufigkeit oder des Umfangs von Datenabfragen kaum eine Aussage darüber treffen, inwieweit dem Gebot der Datenminimierung genüge getan worden ist. Selbst eine einzige Datenabfrage geringen Umfangs kann dem Gebot der Datenminimierung zuwiderlaufen oder eine Vielzahl umfangreicher Datenabfragen

kann mit ihm zu vereinbaren sein. Für das Anwendungsszenario ergeben sich folgende, zu kombinierende **Maßnahmen**:

1. Anlasslose Standortbestimmungen sind zu unterlassen. Hierzu sind zwei Maßnahmen umzusetzen: Standortbestimmungen dürfen (a) nur im Fall einer Standortabfrage und (b) nur bei konkretem Anlass stattfinden.
2. Es dürfen nur Standorte für den konkreten Einsatz in Betracht kommender Beschäftigter bestimmt werden.

Für eine Metrik zur Kontrolle der Datenminimierung bei Standortabfragen sind zur Ermittlung des Umfangs der Datenverarbeitung als **Indikatoren** insb. Metadaten von Abfragen (Zeitstempel, Spezifikation, z. B. durch Filter wie Qualifikation und Zeitraum des Einsatzes), Metadaten ermittelter Standorte wie insb. die Zahl der Betroffenen sowie Daten zu den Betroffenen heranzuziehen. Daneben lässt sich aus Personallisten, Berufsbezeichnungen und Qualifikationen sowie Dienstplänen und Abwesenheitsvermerken schließen, welche Beschäftigten wann für welche Einsätze in Betracht kommen. Ggfs. erlauben die Betriebsabläufe auch die Erhebung von Daten, die indizieren, ob Anlass zur Standortdatenerhebung besteht, z. B. durch in Anlagen verbaute Sensoren.

Um zu überprüfen, dass Standorte nur auf Abfragen hin ermittelt werden, schlagen wir folgende **Metrik 1a** vor, die für einen bestimmten Zeitraum den Anteil durch Abfragen veranlasster Standortbestimmungen (*AvS(Zeitraum)*) bestimmt:

$$\text{AvS(Zeitraum)} = \frac{\text{Zahl Standortabfragen}}{\text{Zahl Standortbestimmungen}} \tag{1a}$$

Wir verwenden den Begriff der Standortbestimmung für die Berechnung eines oder mehrerer Standorte in einem Zusammenhang, wobei wir durch dieselbe Abfrage veranlasste Standortberechnungen als eine Standortbestimmung werten. Ziel ist es, dass die Zahl der Standortabfragen der Zahl der Standortbestimmungen entspricht, Zielwert ist dementsprechend 1. Unter 1 liegende Werte sagen aus, dass mehr Standortbestimmungen stattgefunden haben als aufgrund von Abfragen geboten. Je weiter sich die Kennzahl 0 annähert, umso mehr Standortbestimmungen haben ohne Anfrage und damit anlasslos stattgefunden.

Die Unterlassung anlassloser Standortabfragen lässt sich durch **Metrik 1b** verifizieren, wenn Daten über Anlagenausfälle und erhöhten Personalbedarf verfügbar sind. Metrik 1b bestimmt für einen gegebenen Zeitraum den Anteil der real veranlassten Abfragen für eine standortbasierte Einsatzplanung von der Gesamtzahl der Standortabfragen (*ArA(Zeitraum)*), Zielwert ist wiederum 1:

$$\text{ArA(Zeitraum)} = \frac{\text{Zahl Anlässe}}{\text{Zahl Standortabfragen}} \tag{1b}$$

Die Unterlassung der Erhebung von Standorten Beschäftigter, die nicht für einen Einsatz in Frage kommen, lässt sich durch folgende **Metrik 2** verifizieren, die für eine bestimmte Abfrage die Anzahl der ermittelten Standorte zur Anzahl der für einen Einsatz in Betracht kommenden Beschäftigten setzt (Anteil anlassbezogener Standorte *AaS*), Zielwert ist wiederum 1:

$$\text{AaS} = \frac{\text{Zahl in Betracht kommender Beschäftigten}}{\text{Zahl ermittelter Standorte}} \tag{2}$$

4 Beispielhafte Umsetzung der Metriken

Eine beispielhafte Architektur zur Umsetzung des Metrikensystems veranschaulicht Abb. 1. Das Metrikensystem ist eine unabhängige und vertrauenswürdige Komponente innerhalb des Unternehmensnetzwerks mit klar definierten und beschränkten Zugriffsrechten und Benutzungsschnittstellen.

Die Beschäftigten tragen aktive Wearables am Körper und diese funken ihre technischen Kennungen an die in den Werkshallen verteilten Empfänger, die die gemessenen Signalübertragungszeiten und Kennungen an ein Gateway senden. Das Gateway berechnet in Echtzeit aus den Signallaufzeiten die Positionsdaten der Wearables und damit der Beschäftigten. Die Standorte werden vom PEP-Tool abgefragt und den Beschäftigten zugeordnet. Das Gateway protokolliert die Häufigkeit der Standortbestimmungen, die Zahl der Standortabfragen sowie die Zahl der gemessenen Standorte mit den zugehörigen Zeitpunkten und hält die Daten für einen gewissen Zeitraum vor. Im PEP-Tool werden die Anlässe für die Umplanung von Beschäftigten wie auch die jeweils in Betracht kommenden Beschäftigten hinterlegt. Dazu müssen im PEP-Tool Informationen über Maschinenausfälle, erhöhten Personalbedarf und Katastrophenfälle automatisch durch andere Systeme oder manuell eingetragen werden.

Das Metrikensytem ruft regelmäßig diese Daten vom Gateway und dem PEP-Tool ab, um die Metriken zu berechnen. Die Endgeräte, Infrastrukturkomponenten und weiteren technischen Komponenten müssen für ihre Nutzung im Rahmen des Metrikensystems zwar bestimmte Funktionen und Schnittstellen unterstützen, sind aber nicht auf herstellerspezifische Produkte beschränkt. Anhand der Metriken können z. B. der Betriebsrat, Datenschutzbeauftragte oder auch die Betroffenen selbst den Beschäftigtendatenschutz innerhalb des Unternehmens überwachen.

5 Diskussion und Ausblick

Die *Metriken sind weiterzuentwickeln und zu evaluieren*. Die vorgeschlagenen Metriken beschränken sich auf die Bestimmung von Anteilen. Um ihre Aussagekraft zu erhöhen, ist die Einbeziehung weiterer Daten, wie Zeitstempel und Pseudonyme in Betracht zu ziehen.

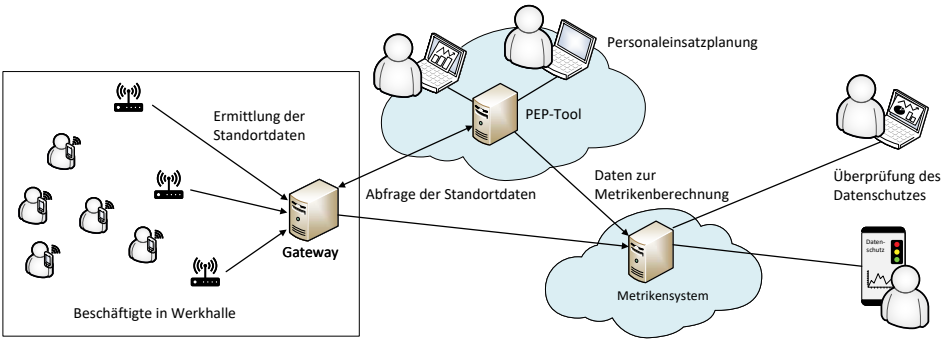


Abb. 1: Beispiel-Architektur des Metrikensystems

Notwendig ist die *Entwicklung weiterer Metriken* zur Kontrolle der Datenminimierung, etwa für andere Anwendungsszenarien. Auch ließe sich eine komplexere Metrik formulieren, mittels derer sich mehrere PEP-Lösungen unter Datensparsamkeits- und Effizienzgesichtspunkten miteinander vergleichen lassen. Dabei wäre der Umfang der mit der PEP-Lösung verbundenen Datenverarbeitungsvorgänge in Beziehung zu den durch sie zu vermeidenden Fehlern und Schwächen der Personalplanung zu setzen. Hierdurch ließe sich validieren, ob die Nutzung von Standortdaten für die agile Einsatzplanung für ein Unternehmen überhaupt erforderlich ist. Weiterhin bedarf es für einen starken, verifizierbaren Beschäftigtendatenschutz weiterer Datenschutzmetriken zur Kontrolle auch anderer Anforderungen des Datenschutzrechts wie etwa Datensicherheit und Speicherbegrenzung.

Literaturverzeichnis

[1] Ammann, F.-E.; Sowa, A.: „Systematische Entwicklung von Metriken zur Beurteilung der Datensicherheit“. Datenschutz und Datensicherheit (DuD) 4/2012, 247–252.

[2] Böhme, R.; Freiling , F. C.: „On Metrics and Measurements“. In: Eusgeld, I.; Freiling, F. C.; Reussner, R. (Hrsg.). „Dependability Metrics“. 2008.

[3] Diel, S.; Kohn, M.; Schleper, J.; Selzer, A.: „Datenschutzmetriken im Beschäftigungsverhältnis“. DuD 12/2021.

[4] Jäger, B.; Selzer, A.; Waldmann. U.: „Die automatisierte Messung von Cloud-Verarbeitungsstandorten“. DuD 1/2015.

[5] Kühling, J.; Buchner, B. (Hrsg.): „DS-GVO / BDSG“. 3. Aufl. 2020.

[6] Paal, B; Pauly, D. A. (Hrsg.): „DS-GVO BDSG“. 3. Aufl. 2021.

[7] Simitis, S.; Hornung, G.; Spiecker gen. Döhmann, I. (Hrsg.): „Datenschutzrecht“. 2019.

[8] Wolff, H. A.; Brink, S.: „BeckOK Datenschutzrecht“. 39. Edition, Stand: 01.02.2022.

Rechtliche Implikationen von Cybersicherheitsvorfällen in der kommunalen Verwaltung

Linda Schreiber[1]

Abstract: Einrichtungen der kommunalen Verwaltung sind zunehmend von IT-Sicherheitsvorfällen betroffen, mit zum Teil weitreichenden Auswirkungen. Diese Vorfälle können sich auf die Verfügbarkeit von IT-Systemen und damit wichtiger kommunaler Leistungen im Bereich der Daseinsvorsorge auswirken sowie auf die Vertraulichkeit und Integrität von Daten und Systemen der kommunalen Verwaltung. Der Beitrag skizziert Beispiele von Cybersicherheitsvorfällen im kommunalen Bereich sowie Anknüpfungspunkte für rechtliche Implikationen und identifiziert weitere Forschungsfragen.

Keywords: Cybersicherheit, öffentliche Verwaltung, Datenschutz, Regulierung, Geheimschutz, Kommunen, Verfassungsrecht

1 Die Rolle kommunaler Einrichtungen

Kommunale Einrichtungen nehmen eine Vielzahl öffentlicher Aufgaben wahr, also Aufgaben „an deren Erfüllung ein gesteigertes Interesse der Gemeinschaft besteht“ [Bu74]. Im Rahmen der kommunalen Selbstverwaltung regeln Kommunen alle Angelegenheiten ihrer örtlichen Gemeinschaft in eigener Verantwortung (Art. 28 Abs. 2 GG). Neben den Aufgaben der kommunalen Daseinsvorsorge erfüllen Kommunen ihnen durch Gesetze übertragene Aufgaben, wie beispielsweise bei der Durchführung von Bundestagswahlen (§9 BWahlG), im Bereich der Sozialhilfe (§ 6 SGB II) oder des Baurechts (§ 1 BauGB). Für BürgerInnen, ansässige Unternehmen und den Staat besteht demnach vielfach eine besondere Abhängigkeit von der Erbringung der Aufgaben und Verfügbarkeit der Dienstleistungen der kommunalen Verwaltung. Besorgniserregend ist vor diesem Hintergrund, dass sich Berichte von Cybersicherheitsvorfällen in kommunalen Einrichtungen häufen.

Der Beitrag zeigt Beispiele für verschiedene Arten von Cybersicherheitsvorfällen in kommunalen Einrichtungen sowie deren Auswirkungen. Zudem werden rechtliche Rahmenbedingungen, die im kommunalen Bereich für die Prävention von und den

[1] Fraunhofer-Institut für Sichere Informationstechnologie SIT | Nationales Forschungszentrum für angewandte Cybersicherheit ATHENE, Rheinstr. 75, 64295 Darmstadt, Deutschland, linda.schreiber@sit.fraunhofer.de.
Diese Forschungsarbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF) und vom Hessischen Ministerium für Wissenschaft und Kunst (HMWK) im Rahmen ihrer gemeinsamen Förderung für das Nationale Forschungszentrum für angewandte Cybersicherheit ATHENE unterstützt.

doi:10.18420/inf2022_22

Umgang mit Cybersicherheitsvorfällen anwendbar sind, skizziert, wobei ein weiterer Forschungsbedarf identifiziert wird.

2 Cybersicherheitsvorfälle in der kommunalen Verwaltung

2.1 Definition Cybersicherheitsvorfälle und Schutzziele

Der Begriff der Cybersicherheit und der damit verbundene Schutz von Informationen und Systemen bestimmt sich über die Bewahrung beziehungsweise die Einhaltung der Schutzziele der Vertraulichkeit, Integrität und Verfügbarkeit von Informationen. Dementsprechend sind *Cybersicherheitsvorfälle* ungewollte oder unerwartete Ereignisse, wie Angriffe oder unautorisierte Zugriffe, die nachteilige Auswirkungen haben, in dem sie eines oder mehrere ebendieser Schutzziele beeinträchtigen (Art. 4 Nr. 2 und 7 NIS Richtline). Die *Vertraulichkeit* von Informationen und Systemen bedeutet, dass nur ein jeweils definierter Kreis an befugten Personen Zugriff auf diese haben bzw. Informationen nur gegenüber diesen offengelegt werden. Die *Verfügbarkeit* von Informationen und IT-Systemen bezieht sich auf deren Eigenschaft, dass sie wie vorgesehen zugänglich und nutzbar bzw. betriebsbereit sind. *Integrität* bezieht sich auf die Korrektheit und Unverfälschtheit von Daten bzw. die korrekte Funktionsweise von IT-Systemen, also den Schutz vor einer unbefugten Manipulation. Daneben gibt es noch eine Reihe weiterer, untergeordneter Schutzziele bzw. Spezialfälle der Dreien, wie die *Authentizität*, *Unverkettbarkeit*, oder *Nachweisbarkeit* [BA10]. Viele Cybersicherheitsvorfälle sind gezielte Angriffe, also eine vorsätzliche Gefährdung der Cybersicherheit durch einen Angreifer mit dem Ziel, sich oder einem anderen Vorteile zu verschaffen [BS22].

2.2 Beispiele zu Cybersicherheitsvorfällen auf kommunaler Ebene

In Einrichtungen der kommunalen Verwaltung haben sich in den vergangenen Jahren zahlreiche Cybersicherheitsvorfälle unterschiedlichen Ausmaßes ereignet. Im Folgenden wird anhand einer Auswahl von Beispielen aus der Praxis die Bandbreite an Cybersicherheitsvorfällen und Beeinträchtigungen der jeweiligen Schutzziele aufgezeigt sowie die praktischen Auswirkungen.

Ransomware Angriffe

Eine häufige Form von Schadsoftware, die von Angreifern bei einer Vielzahl von Cybersicherheitsvorfällen im öffentlichen Bereich verwendet wird, ist Ransomware. Bei Ransomwareangriffen werden Daten und Systeme von Anwendern verschlüsselt und somit der Zugriff verhindert. Angreifer fordern die Zahlung eines Lösegelds, damit die Daten wieder entschlüsselt werden. Betroffen sind bei Kommunen häufig in umfassender Weise verschiedene Ämter, Bereiche und Standorte. Insbesondere wenn Backups der Systeme nicht vorhanden sind, ebenfalls verschlüsselt wurden, veraltet oder in Folge des

Angriffs nicht mehr vertrauenswürdig sind, kann die Wiederherstellung bis zu mehreren Monaten dauern [SK20] [BS22]. Ransomware richtet sich in erster Linie gegen das Schutzziel der Verfügbarkeit. Die Nicht-Verfügbarkeit von Daten und IT-Anwendungen auf technischer Ebene führt dazu, dass viele öffentliche Aufgaben und Dienste, die darauf basieren, nicht erbracht werden können. Dies führt beispielsweise dazu, dass allgemein die Kommunikation mit den verschiedenen Bereichen der öffentlichen Verwaltung eingeschränkt ist, aber auch dazu, dass beispielsweise Sozialhilfe- oder Unterhaltszahlungen nicht erfolgen, neue Kraftfahrzeuge nicht zugelassen oder Personalausweise und Pässe nicht ausgestellt werden [St21] [Sc21].

Neben der Verfügbarkeit kann es auch zu Beeinträchtigungen des Schutzziels der Vertraulichkeit kommen, wenn es während des Angriffs gleichzeitig zum Abfluss von Daten kommt. So gibt es Fälle, bei denen personenbezogene Daten von BürgerInnen, VerwaltungsmitarbeiterInnen und Abgeordneten nach einem Ransomware-Angriff im Darknet veröffentlicht wurden [He21] [St21].

Distributed Denial of Service

„Denial of Service“ (DoS) Angriffe sind Angriffe, die zu einer gezielten Überlastung von Diensten, Systemen oder Webseiten führen und sich somit gegen das Schutzziel der Verfügbarkeit richten. Dies geschieht häufig in der Form von verteilten “Distributed Denial of Service” (DDoS) Angriffen, bei denen der Angriff durch Anfragen von vielen Rechnern gleichzeitig erfolgt. Häufig haben Angreifer hierfür zuvor die Kontrolle über eine Vielzahl an Rechnern im Netz mittels Schadsoftware übernommen und damit ein sog. Botnetz erstellt, womit ein entsprechender Angriff durchgeführt wird [GW21] [SK20]. Im öffentlichen Bereich konnte man diese Angriffsart zuletzt insbesondere gegen Schul- und Lernplattformen beobachten, aber bspw. auch gegen ein Portal zur Vergabe von Impfterminen [Sä20] [Sp21] [BS21]. DDoS-Angriffe führen dazu, dass betroffene Systeme und Dienste zeitweise nicht oder nur eingeschränkt für die Nutzenden erreichbar sind.

Phishing

Phishing ist eine Form des Social Engineerings und beschreibt eine Methode, bei der mittels gefälschter bzw. manipulierter E-Mails oder Webseiten, in denen sich der Angreifer als vertrauensvolle, bekannte Quelle ausgibt, Nutzer zu einer bestimmten Handlung verleitet werden. Dies kann etwa die Eingabe von Daten und Passwörtern sein, oder das Öffnen eines Anhangs, was zur unbemerkten Ausführung eines Schadprogramms führt [We22] [BS22]. Bei Phishing-Vorfällen, die öffentliche Stellen involvieren, sind diese nicht immer das Opfer: In Nordrhein-Westfalen haben Betrüger das Portal für Corona-Nothilfen des Wirtschaftsministeriums kopiert und darüber Daten von Selbstständigen und Kleinunternehmen erlangt, womit sie wiederum die Nothilfen selbst beantragt haben [Ko20]. Phishing-Angriffe, die sich gegen öffentliche Einrichtungen selbst richten, können unterschiedlichste Auswirkungen haben. Sie können als Angriffsvektor zur Ausführung von Malware genutzt werden und damit beispielsweise zu oben genannten Ransomware-Angriffen führen [SN21].

Technische Schwachstellen

Neben gezielten Cyberangriffen kommt es auch zu Cybersicherheitsvorfällen, die nicht durch Angreifer vorsätzlich initiiert wurden, sondern durch technisch bedingte Schwachstellen, die durch Angreifer ausgenutzt werden können. Technische Schwachstellen sind sicherheitsrelevante Fehler in IT-Systemen, die das jeweilige System bzw. die einsetzende Einrichtung anfällig für Bedrohungen und Cyberangriffe machen [BS22]. In diesem Zusammenhang wird auch der Begriff der Sicherheitslücke für die Eigenschaft von IT-Systemen verwendet, bei denen es durch Ausnutzung möglich ist, dass sich Dritte gegen den Willen des Berechtigten Zugang zum jeweiligen System verschaffen oder die Funktion des jeweiligen Systems beeinflussen können (§ 2 Abs. 6 BSIG). Sicherheitslücken wurden im öffentlichen Bereich bspw. in einer App für Rettungskräfte sowie in Software zur Wahlauswertung entdeckt. Hierbei verletzten die Sicherheitslücken das Schutzziel der Vertraulichkeit, durch die Möglichkeit der Einsichtnahme von Patienten- und Einsatzdaten, bzw. das Schutzziel der Integrität, durch die Möglichkeit der Manipulation von Wahlergebnissen [Do18] [Go20].

3 Rechtliche Implikationen

Es zeigt sich, dass die Auswirkungen von Cybersicherheitsvorfällen in kommunalen Einrichtungen unterschiedlichster Art und Tragweite sein können. Bestehende Handreichungen, wie beispielsweise das „IT-Grundschutz-Profil Basis-Absicherung Kommunalverwaltung“ [DD22] und die „Handreichung zur Ausgestaltung der Informationssicherheitsleitlinie in Kommunalverwaltungen“ [Ko17] bieten für das spezifische Umfeld von Kommunen praktische Hilfestellung zur Implementierung technischer und organisatorischer Maßnahmen sowie der Errichtung eines Informationssicherheits-Managementsystems. Allerdings fehlt es bislang an einer umfassenden systematischen Analyse der rechtlichen Rahmenbedingungen, die im kommunalen Bereich für die Prävention von und den Umgang mit Cybersicherheitsvorfällen anwendbar sind und welcher cybersicherheitsrechtliche Handlungsbedarf sich aus diesen ergibt. Mögliche Anknüpfungspunkte für künftige Forschung sind die folgenden Bereiche:

Datenschutzrecht

In vielen Fällen führen Cybersicherheitsvorfälle zu Datenschutzverletzungen (Art. 4 Nr. 12 DSGVO). Durch die verschiedenen Einrichtungen der kommunalen Verwaltung werden vielfach personenbezogene Daten in einem besonderen Umfang, Detailgrad, Bandbreite und Schutzbedarf verarbeitet.

Zudem gelten im kommunalen Bereich zum Teil bereichsspezifische datenschutzrechtliche Regelungen, wie beispielsweise zum Steuergeheimnis für Amtsträger in Finanzbehörden (§ 30 AO) oder dem Sozialgeheimnis für Sozialleistungsträger (§ 35 SGB I, § 67 SGB X).

Hier gilt es beispielsweise spezifische Kriterien für die bei Datenschutzverletzungen durchzuführende Risikobewertung zu analysieren [Ar18] [Eu21]. Für die Feststellung der Betroffenheit von Personen sowie der Bewertung des sich für diese jeweils aus der Datenschutzverletzung ergebenden Risikos wird in diesen Fällen oftmals ein kontinuierliches Monitoring über längere Zeiträume notwendig sein. Dies ist insbesondere bei komplexen, unübersichtlichen Cybersicherheitsvorfällen wie Ransomwareangriffen der Fall, wenn zum Beispiel nicht klar ist, ob und welche Daten während eines Angriffs abgeflossen sind und wozu die Angreifer diese verwenden.

Geheimschutz

Zudem ergibt sich ein besonderer Schutzbedarf für Verschlusssachen nachdem Sicherheitsüberprüfungsgesetz (SÜG) sowie den entsprechenden landesrechtlichen Regelungen. Verschlusssachen sind im öffentlichen Interesse, insbesondere zum Schutz des Wohles des Bundes oder eines Landes, geheimhaltungsbedürftigen Tatsachen, Gegenstände oder Erkenntnisse (§ 4 Abs. 1 S. 1 SÜG). Verschlusssachen werden entsprechend ihrer Schutzbedürftigkeit, also dem Schaden den eine unbefugte Kenntnisnahme erzeugen kann, eingestuft (§ 4 Abs. 2 SÜG).

Onlinezugangsgesetz

Durch das Onlinezugangsgesetz (OZG) werden Bund, Länder und Kommunen dazu verpflichtet bestimmte Verwaltungsleistungen auch digital über Verwaltungsportale anzubieten, die Verwaltungsportale werden zu einem Portalverbund verknüpft (§ 1 OZG). Für die IT-Sicherheit der an den Portalverbund angebundenen IT-Komponenten sind verbindliche Maßnahmen und Standards durch die IT-Sicherheitsverordnung Portalverbund (ITSiV-PV) definiert. Dazu zählen beispielsweise Technische Richtlinien, die durch das BSI herausgegeben werden sowie verbindliche Penetrationstests und Webchecks (§ 2 Abs. 2 und 6 ITSiV-PV). Kritik im Zusammenhang mit der IT-Sicherheitsverordnung Portalverbund gab es in Bezug auf die späte Veröffentlichung des Standards zu einem Zeitpunkt, zu dem entsprechende Software bereits entwickelt war und Anforderungen nachträglich umzusetzen sind. Zudem sehen Kommunen Probleme bei der konkreten Umsetzung der Richtlinie sowie dem entstehenden finanziellen Aufwand [Pu22].

BSI-Gesetz und Richtlinie zur Netz- und Informationssicherheit

Zum Teil sind kommunale Einrichtungen der Kritischen Infrastruktur (KRITIS) nach § 2 Abs. 10 BSIG zu zuordnen, sofern bspw. kommunale Eigenbetriebe Aufgaben im Bereich der KRITIS Sektoren, wie Wasser oder Energie übernehmen und die entsprechenden Schwellenwerte erreichen (Verordnung zur Bestimmung Kritischer Infrastrukturen nach dem BSI-Gesetz). Für Betreiber Kritischer Infrastrukturen gelten besondere Anforderungen an die technischen und organisatorischen Vorkehrungen zur Vermeidung von Cybersicherheitsvorfällen sowie zur Meldung von Störungen (§§ 8a ff. BSIG).

Auf europäischer Ebene definiert die Richtlinie (EU) 2016/1148 zur Gewährleistung einer hohen Netzwerk- und Informationssicherheit (NIS-Richtlinie) einheitliche Anforderungen an die Mitgliedsstaaten sowie Betreiber Kritischer Infrastrukturen. Die Europäische Kommission hatte im Dezember 2020 einen Vorschlag für eine überarbeitete NIS-Richtlinie (NIS-2-Richtlinie) vorgelegt [Eu20]. Im Mai 2022 wurde eine vorläufige, bislang unveröffentlichte Einigung erzielt, die noch angenommen werden muss. Grundsätzlich sieht die NIS-2-Richtlinie einen wesentlich breiteren Anwendungsbereich vor. Ein wesentlicher Diskussionspunkt in den Verhandlungen war die Frage, welche Ebenen der öffentlichen Verwaltung von den Regelungen erfasst werden sollen. Die Einigung sieht eine grundsätzliche Einbeziehung der öffentlichen Verwaltung vor, räumt den Mitgliedsstaaten aber Spielraum bei der Beurteilung ein, inwiefern Einrichtungen auf lokaler Ebene erfasst werden [Be22] [Rh22]. Ein wichtiger Gesichtspunkt künftiger Forschungsarbeiten wird die Analyse der NIS-2-Regelungen und der Umsetzung in Deutschland sein sowie die Auswirkungen hiervon im spezifischen Umfeld der föderalen Verwaltungsstrukturen in Deutschland.

Verfassungsrechtliche Anforderungen

Ein Forschungsbedarf lässt sich zudem bei der Untersuchung verfassungsrechtlicher Anforderungen an die Cybersicherheit der kommunalen Verwaltung erkennen sowie bei der Frage, ob bestehende und künftige gesetzliche Regelungen diesen Anforderungen entsprechen. Im Hinblick auf die zunehmende Bedeutung der Nutzung informationstechnischer Systeme hat das Bundesverfassungsgericht 2008 erstmals ein grundrechtlich erhebliches Schutzbedürfnis festgestellt und das Grundrecht auf Gewährleistung der Vertraulichkeit und Integrität von IT-Systemen als Ausprägung des allgemeinen Persönlichkeitsrechts beschrieben [Bu08]. Der Gewährleistungsgehalt von Grundrechten beschränkt sich nicht auf deren Abwehrfunktion, sondern kann auch staatliche Schutzpflichten begründen, also eine konkrete Pflicht des Staates dazu beizutragen, dass die Cybersicherheit von IT-Systemen geschützt wird [Bu21].

4 Ausblick

In Anbetracht des hohen Maßes an gesellschaftlicher, wirtschaftlicher und staatsorganisatorischer Abhängigkeit von der Verfügbarkeit kommunaler Leistungen sowie der Vertraulichkeit von Daten im kommunalen Wirkbereich wird die Untersuchung der rechtlichen Rahmenbedingungen und Anforderungen an die Cybersicherheit im spezifischen Umfeld der kommunalen Verwaltung in Zukunft wichtiger Forschungsgegenstand sein. Die skizzierten rechtlichen Anknüpfungspunkte sind nicht abschließend und lassen sich insbesondere hinsichtlich konkreter Handlungsbedarfe sowie ihrer Angemessenheit weiter untersuchen. Dabei sind insbesondere auch praktische Herausforderungen im Kontext der kommunalen Selbstverwaltung, der Organisationsstrukturen sowie der finanziellen und personellen Umsetzbarkeit von Anforderungen zu analysieren.

5 Literaturverzeichnis

[Ar18] Artikel-29-Datenschutzgruppe, Leitlinien für die Meldung von Verletzungen des Schutzes personenbezogener Daten, zuletzt überarbeitet und angenommen am 06.02.2018.

[BA10] Bedner, M.; Ackermann, T.: Schutzziele der IT-Sicherheit, DuD 2010, S. 323-328.

[Be22] Bertuzzi, Luca: NIS-2 kommt – und nun?, Tagesspiegel Background, https://background.tagesspiegel.de/cybersecurity/nis-2-kommt-und-nun, 16.05.2022.

[BS21] Bundesamt für Sicherheit in der Informationstechnik, Die Lage der IT-Sicherheit in Deutschland 2021

[BS22] Bundesamt für Sicherheit in der Informationstechnik, Glossar der Cybersicherheit, https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Glossar-der-Cyber-Sicherheit/glossar-der-cyber-sicherheit_node.html, Stand: 28.04.2022.

[Bu21] Bundesverfassungsgericht, Beschluss vom 08.06.2021, 1 BvR 2771/18.

[Bu08] Bundesverfassungsgericht, Urteil vom 27.02.2008, 1 BvR 370/07 u. 595/07.

[Bu74] Bundesverfassungsgericht, Beschluss vom 18.12.1974, 1 BvR 430/65 u. 259/66.

[DD22] Deutscher Städtetag, Deutscher Landkreistag, Deutscher Städte- und Gemeindebund: IT-Grundschutz-Profil Basis-Absicherung Kommunalverwaltung, 31.03.2022.

[Do18] Does S.: Security-Notruf – Sicherheitslücke in Notfall-App legt Einsatzdaten offen, c't 7/2018, S.60.

[Eu20] European Commission, Proposal for a Directive on measures for a high common level of cybersecurity across the Union, repealing Directive (EU) 2016/1148, 16.12.2020.

[Eu21] Europäischer Datenschutzausschuss, Guidelines 01/2021 on Examples regarding Personal Data Breach Notification, Version 2.0, Stand 14.12.2021.

[Go20] Golem, Schwere Sicherheitslücken in bayerischer Wahlsoftware, https://www.golem.de/news/ok-vote-schwere-sicherheitsluecken-in-bayerischer-wahlsoftware-2012-153062.html, Stand: 28.04.2022.

[GW21] Grimm, R.; Waidner. M.: IT-Sicherheit aus technischer Sicht. In (Hornung, G.; Schallbruch, M., Hrsg): IT-Sicherheitsrecht Praxishandbuch. Nomos, Baden-Baden, 1.Auflage, S. 33-62, 2021.

[He21] Heise online, Witten: Bei Cyberangriff erbeutete Daten im Darknet veröffentlicht, https://www.heise.de/news/Hacker-veroeffentlichen-Wittener-Daten-Buergermeister-warnt-6269952.html, Stand: 28.04.2022.

[Ko20] Kommunal, So schützen sich Behörden von Cyber-Kriminellen, https://kommunal.de/schuetzen-vor-cyber-kriminellen.

[Ko17] Kommunale Spitzenverbänden, Handreichung zur Ausgestaltung der Informationssicherheitsleitlinie in Kommunalverwaltungen, Februar 2017.

[Pu22] Punz, Matthias: OZG: Kritik an Verordnung zu IT-Sicherheit, Tagesspiegel Background, https://background.tagesspiegel.de/smart-city/ozg-kritik-an-verordnung-zu-it-sicherheit, 25.01.2022.

[Rh22] Rheinland-Pfalz Landesvertretung Brüssel, Rat und Parlament erzielen Einigung über EU-Cybersicherheits-niveau, https://europa.rlp.de/de/aktuelles/detail/news/News/detail/rat-und-parlament-erzielen-einigung-ueber-eu-cybersicherheits-niveau/, 20.05.2022.

[Sä20] Sächsisches Staatsministerium für Kultus, Cyberangriff legt LernSax lahm, https://www.bildung.sachsen.de/blog/index.php/2020/12/15/cyberangriff-legt-lernsax-lahm/.

[Sc21] Landeshauptstadt Schwerin, IT-Störungsmeldung, https://www.schwerin.de/it-stoerungsmeldung/, Stand: 28.04.2022.

[SK20] Sohr, K.; Kemmerich, T.: Technische Grundlagen der IT-Sicherheit. In (Kipker, D.-K., Hrsg): Cybersecurity Rechtshandbuch, 1. Auflage, C.H. Beck München, S. 23-82, 2020.

[Sp21] Spiegel Online, Cyberangriff auf Schulplattform, https://www.spiegel.de/netzwelt/web/rheinland-pfalz-14-jaehriger-stoerte-von-schulen-genutztes-konferenzsystem-a-664fa36a-c962-42a8-bca6-529342b81c79.

[St21] Stiftung Neue Verantwortung, Transkript zum Hintergrundgespräch: "Cyberkriminelle erpressen Anhalt-Bitterfeld - Was können wir daraus lernen?", https://www.stiftung-nv.de/de/publikation/transkript-zum-hintergrundgespraech-cyberkriminelle-erpressen-anhalt-bitterfeld-was#collapse-newsletter_banner_bottom, Stand 28.04.2022.

[We22] Weidemann, M.: Lexikon des Strafrechts – Computerkriminalität. In (Heintschel-Heinegg, B., Hrsg): BeckOK StGB, 52. Edition, C.H. Beck, München, 2022.

Künstliche Intelligenz für kleine und mittlere Unternehmen (KI-KMU 2022)

AI-based Online Vacancies - Trends and Differences between SMEs and Larger Enterprises

AI-based Vacancies of SMEs and Large Enterprises

Maximilian Lowin[1], Jendrik Seip[2]

Abstract: Artificial intelligence is an essential key competence for cost reduction in companies. For this reason, the German government is promoting its use and hopes that it will lead to a digital transformation of companies. We investigated whether small and medium-sized enterprises (SMEs) also participate in this transformation by examining AI-related vacancies for their fields of activity. In particular, we observe whether the required skills of potential applicants suggest that enterprises also require artificial intelligence at an advanced stage and that quality factors such as fairness and ethics play a role. We found that the need for more advanced AI-related jobs has started increasing but still plays a minor role compared to traditional data science roles. Notably, only large companies require some of these additional skills. Moreover, many vacancies even state that a deep AI understanding is not necessarily a prerequisite.

Keywords: job advertisement, vacancies, digital transformation, data science, artificial intelligence.

1 Introduction

With its AI strategy, the German government promotes artificial intelligence in the economy. In particular, small and medium-sized enterprises (SMEs) are to be enabled to make the transfer to a data-driven company. The AI strategy has been in place since 2018, and the current federal government continues this strategy. By 2025, 5 billion euros are to be invested in promoting AI as part of a 'future package' [Bu20]. We, therefore, try to measure whether the AI transformation is underway by monitoring job postings and specifying the fields of activity of AI experts.

We are interested in the differences of job vacancies between large enterprises (LEs) and SMEs and follow the EU definition to separate the companies into SMEs and large companies. According to the EU definition an SME must have fewer than 250 employees and less than 50 million € in sales [Eu03]. In order to gain a broad overview of companies in Germany, we use job advertisements from one of the largest online vacancy platforms in Germany. These vacancies mainly state the requirements of the applicants, e.g., whether applicants should only know what artificial intelligence is or, for example, have a more

[1] J. W. Goethe Universität Frankfurt a. M., Professur Wirtschaftsinformatik und Informationswissenschaften, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, lowin@wiwi.uni-frankfurt.de,
[2] J. W. Goethe Universität Frankfurt a. M., Professur Wirtschaftsinformatik und Informationswissenschaften, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, seip@wiwi.uni-frankfurt.de

doi:10.18420/inf2022_23

profound knowledge of the usual programming libraries. Based on this, we want to deduce what the activities of the applicants are and how far the adoption of artificial intelligence in companies has progressed.

2 Theoretical Background and Research Setting

The combination of digitalization and artificial intelligence has the potential to substitute human employment [BD18]. Thus, there is a concern that AI might create mass unemployment due to automation [Be18]. However, humans are still advantageous over computers depending on the individual situation [BD18]: There is not only one type of labor as the work of [ALM03] classifies labor among the degree of cognition and the degree of reputation. This separation indicates that incorporating AI is more beneficial for repetitive and routine tasks than non-routine tasks when a computer can follow a predefined set of rules [BD18]. However, it is essential to note that AI does replace not only specific jobs but also creates new ones. Until today, there is no one-fits-all solution for a universal AI. Thus, software developers need to create complex systems to facilitate the benefits of AI in real-world usage. Furthermore, the need to analyze data is increasing, and the new role of data scientists in teams is becoming increasingly important [Ki18]. Data scientists combine varied skills ranging from business understanding to big data, Machine Learning, Statistics, Math, and Operations Research [Ki18].

However, new problems have occurred with the increasing availability of AI-based solutions. Researchers already identified potential ethical issues two decades ago [Bo03]. While those works relate to “superintelligence” [Bo03], there is also recent literature on weak AI solutions. When AI-based systems learn from biased human decisions, AI advice can have discriminatory effects, e.g., based on the training data or feature selection [Bo18]. Thus, only implementing data science solutions might not be sufficient. In a study by Accenture [WDM17], the authors identified three new categories of human jobs emerging as a consequence of new AI solutions. These jobs consist of trainers, explainers, and sustainers. According to [WDM17], trainers teach AI systems how to perform and make fewer errors, e.g., by letting an AI behave more humanly, have more empathy, or be fairer. They define explainers as a bridge between technology and business leaders, increasing AI-system transparency by explaining how these systems work and how they make specific decisions. Last, sustainers ensure that AI systems operate as designed. For instance, they ensure fairness, ethics, and compliance and try to reduce biases and discrimination [WDM17].

In our study, we want to summarize the current state of job demand five years after the paper of [WDM17] to analyze whether the job market picks up the new job categories and which job roles are prevalent in the current market. Therefore, we use various AI-based vacancies of SMEs and LEs and analyze them according to various job types.

3 Methodology

To shed more light on the different trends and characteristics, we used data from online vacancies, which we will present in the next section first. Afterward, we describe our preprocessing to allow a proper comparison between the different job advertisements. Last, we explain our manual labeling process to classify each vacancy into a preselected job type.

3.1 Data

For our dataset, we used job listings from one of the largest online vacancy platforms in Germany. The platform separates each vacancy into five texts: introduction, description, profile, we offer, and contact. In those job listings, we searched in May 2022 for texts that contained 'artificial intelligence', 'künstliche Intelligenz', 'AI', or 'KI'. We used regular expressions to ensure that 'AI' and 'KI' were not part of the same word, but compound words such as 'KI-Experte' were detected. In the next step, we used the EU definition to separate the companies into SMEs and large companies, according to which an SME must have fewer than 250 employees and less than 50 million € in sales [Eu03]. We matched the companies' names with available Wikipedia articles to automatically determine the number of employees. For this purpose, we use the Wikipedia Python library. In some cases, the company names do not exist on Wikipedia exactly as the vacancy platform provides them, e.g., the platform lists vacancies from the company 'Mercedes-Benz', parent company 'Mercedes-Benz Group AG', Wikipedia article 'Mercedes-Benz Group'. Therefore, we checked the correctness of the search results manually. If no Wikipedia article was available, we searched the company names manually via Google search and collected information about the revenue and the number of employees. In total, our dataset results in 1,589 vacancies from 530 companies. 144 of these companies are SMEs, which posted 243 vacancies in our dataset.

3.2 Preprocessing

The job listings considered in our dataset are mainly in German and partly in English. The German listings were translated to English using Google Translator via the Python package deep_translator to homogenize the data. This step allowed better comparison and was a requirement for later word manipulation software. Since we want to find the most frequent word combinations, we have to clean the texts from stop words and reduce them to their root words (stemming) for the n-grams ($n \in \{1,2,3\}$). Thus, 'intelligence' and 'intelligent' are assigned to the same group. We stem the data with the Porter-Stemmer algorithm from the Natural Language Toolkit Python library (NLTK). The stop words database came from NLTK Corpus and the Python library stop_words.

We further used the n-grams for trend detection in word clouds. Although the word endings have been removed by stemming, they are displayed again in the word cloud to facilitate readability.

Since our dataset contained postings that do not relate to AI, we filtered in a last preprocessing step all postings that did not contain at least one of the following keywords in the *profile* or *description* text: AI, artificial intelligence, ML, machine learning, data science, and neural network. This step resulted in 935 vacancies remaining in our dataset.

3.3 Labeling

Consequently, we manually categorized all job postings into one of the following seven job categories:

None: A worker who has no contact with artificial intelligence. The term 'artificial intelligence' might appear in the company description as a buzzword, or the employee has coworkers who will have contact with AI.

Supplemental AI knowledge: It would be an advantage if the applicant has background artificial intelligence knowledge. However, it is not a mandatory requirement, and an applicant may compensate for it with other areas of knowledge.

Data processor: An employee who collects data and puts it into a neat form so that a coworker can analyze the data with artificial intelligence.

Data scientist: A mathematically literate employee who can work with easily usable AI solutions.

Trainer: A developer who wants to improve artificial intelligence so that the AI makes fewer mistakes, is fairer, is more human, and improves itself.

Explainer: An employee who wants to improve the explainability of the AI. The explainer should help transform black-box models into transparent smart decisions and explainable strategies. Explainers are someone who builds trust with, e.g., **L**ocal **I**nterpretable **M**odel-Agnostic **E**xplanations (LIME) or **Sh**apley **A**dditive Ex**p**lanation (SHAP). An explainer analyses AI or creates bridges to corporate communication.

Sustainer: An employee who tries to avoid unintended consequences of the AI. Sustainers enhance fairness, compliance, ethics, scalability, and performance and reduce cost, discrimination and bias.

For the labeling, we manually screened through all vacancies in the dataset in a self-constructed labeling software. To support the labeling process, the labeling software highlighted the keywords used during the filtering process in the preprocessing step in the vacancies to help find the relevant sections of texts. Since the categories listed above do intersect with each other, we classified the ones that fit the best. For the classification

decision, we mainly use the texts from the introduction, description and profile. To ensure inter-rater reliability, we labeled the first 150 vacancies alternately together, similar to pair programming. Our labeling procedure resulted in the following distribution of labels among SMEs and LEs shown in Fig. 1:

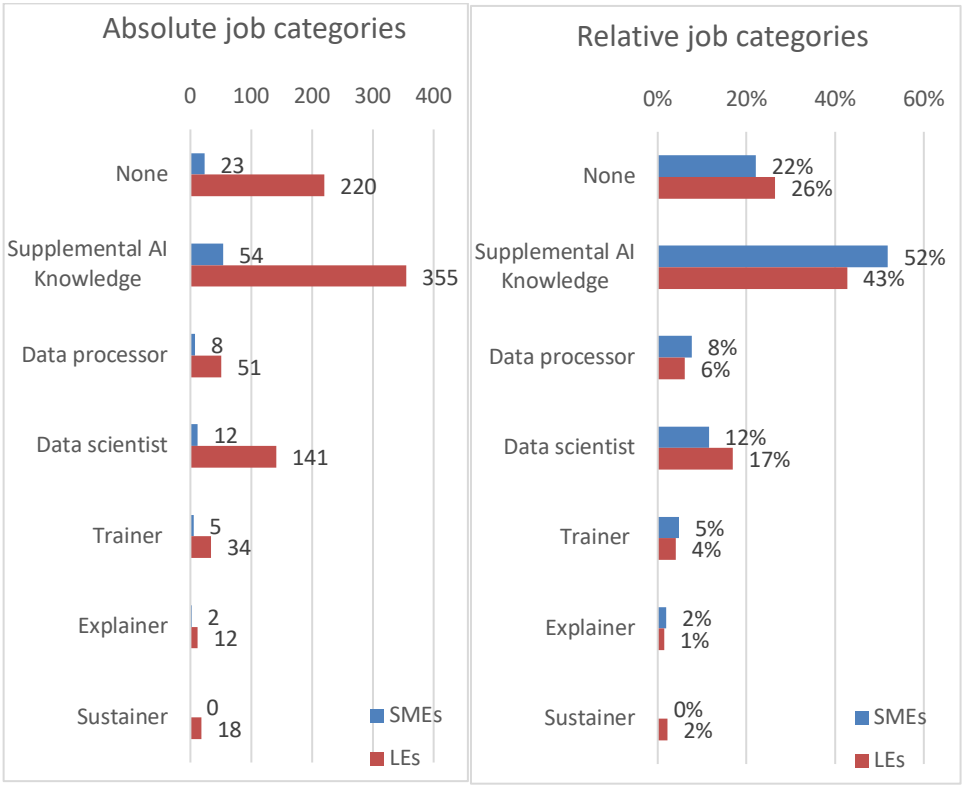


Fig. 1: Absolute and Relative Share of Job Categories

4 Results

Based on our preprocessing and labeling, we analyzed the different n-grams among the text categories offered by the online vacancy platform. Next, we analyzed the different texts among the company classification into SMEs and LEs and the vacancies' label dimensions.

4.1 General Findings

At first glance, all five categories seem to be interconnected since they all relate to the general topics of AI and technology. However, we can identify specific word and word combinations from the respective categories. The *introduction* mainly focuses on the topics of technology, development, and digitalization, which are among the most frequently used unigrams. Furthermore, companies see AI as a future topic and a current trend. The respective bigrams support this view since also 'shape futur[e]' and 'digit[al] transform[ation]' are among the most used bigrams. Besides the technological view, the *introduction* also frequently consists of the words 'team' and 'customer', which are essential stakeholders in a company's everyday business.

The second category *description* focuses on development, data, and the project, see Fig. 2. Again, the terms ‘team’, and ‘custom[er]’ appear frequently. Although, the bigrams of the *description* now mainly include AI-based keywords (‘artifici[al] intellig[ence’, ‘machin[e] learn[ing]’, or ‘big data’), most n-grams overlap with the *introduction*.

In contrast, the category profile includes more specific requirements for the potential applicants, ‘experi[ence]’, ‘knwoledg[e]’, and ‘skill’ are the most used unigrams. Additionally, the bigram ‘sever[al] year[s]’ indicates that most vacancies look for more experienced staff and not for young professionals only. ‘complet[ed] studi[es]’, ‘[busi]ness informat[ic]’, ‘data scienc[e]’, ‘master[‘s] degre[e]’ and ‘busi[ness] administr[ation]’ are among the most used bigrams, indicating that companies are mainly looking for applicants with an university degree. Last, the terms ‘English’ and ‘German’ also appear frequently as well as the term ‘python’.

The next dimension *we offer* includes terms like ‘flexibl[e] work[ing]’, ‘mobil[e] work[ing]’, ‘work-lif[e] balance’, and ‘home offic[e]’, highlighting the benefits of jobs affected by digitalization. The terms ‘health’, ‘train[ing]’, ‘childcar[e]’ and ‘pension’ point out further benefits from the respective companies.

Last, the dimension *contact* only includes basic contact information and typical farewell phrases. Therefore, we exclude this dimension from the further analysis.



Fig. 2: Word Cloud of the Category *Description*

4.2 SMEs vs. Large Enterprises

In the next step, we study the differences in the frequent n-grams between SMEs and large enterprises. Starting with the *introduction*, SMEs mainly focus on the team, customer, and solution itself. All unigrams have around the same share among all words. In contrast, LEs focus on the technology, the company, and the work itself. Interestingly, the unigram 'us' is the fourth most frequent unigram. Additionally, LEs highlight the benefits of their respective industry, while SMEs highlight regionality.

While there are quite a few differences between the n-grams in the *introduction*, the *description*s' n-grams are about the same. Most vacancies highlight development, task, customer, project, and the team. Only SMEs indicate the term 'support' more often, while LEs focus on data and processes.

Furthermore, all *profile* texts emphasize the experience of the applicants. SMEs instead focus on English, a degree, and the programming languages Java, SQL, and C. On the contrary, LEs focus on Python and computer science, and the number of mentions of German and English is equal. The bigrams between both types of companies are equal and contain the AI-related buzzwords.

In the last dimension *we offer*, the SMEs mainly focus on the company and the team and emphasize flat hierarchies. LEs, in contrast, focus on mobile working, work-life balance, support, and pension.

4.3 Types of Vacancies

Based on our manual classification of all vacancies among the three dimensions of a trainer, explainer, and sustainer based on [WDM17], we extended the vacancy types by the dimensions of data engineer, data scientist, and supplemental AI knowledge.

Trainer: All *introduction* texts focus on innovation, experience, systems, and ideas. A new term is the unigram research, which indicates that the job type trainer is not part of an everyday data science team yet. When considering the LE-specific terms, we find a special need for trainers in the automotive industry, consumer goods, and energy buildings. Due to the small absolute number of vacancies for SMEs, we cannot find clear trends here. We can also find automotive-related n-grams in the *description* text, i.e., 'autonom[ous'] drive', but also 'signal process[ing]' or 'comput[er] vision'. The *description* also focuses on AI-based terms and the topic of responsibility, but also the facing of challenges, underlining the need for better model training. The profile-related n-grams again focus on knowledge, degrees, and experience. In contrast to the other job types, vacancies for trainers explicitly focus on skills like Python, deep learning, C, neural networks, TensorFlow, or PyTorch. Interestingly, there are fewer specific frameworks in the vacancies for LEs than for SMEs.

Explainer: Vacancies for explainers again focus on software development and researchers. They emphasize communication and development, but also customer management, support, and safety. Especially for SMEs, exploring and exchanging are essential procedures for explainers. The *profile* texts highlight scientific education like computer science or business informatics studies and ask for team players and communication skills. This finding aligns with the definition of explainers as an intermediary between the algorithm and the decision-makers.

Sustainer: As the last job type described by [WDM17], the sustainers' task is to ensure fair and ethical AI and foster compliance in the company. We can also find this focus in the *introduction* and *description* texts of the vacancies since they relate to quality improvements and incorporate further stakeholders (e.g., terms 'external', 'cooperation', or 'related') and emphasize governance and policy. Again, research is also a frequent unigram for this job type. Considering the job profile, we see an equal share between the need for computer scientists, but also business studies and management students who are technically versed and have an analytical skill set. While we crawled vacancies of trainers and explainers for both SMEs and LEs, we did not find any vacancies for sustainers for SMEs.

Data Processor: Besides the three categories identified by [WDM17], we found an increasing number of data engineers in the vacancies related to the trainers since they prepare the data for model training and improvement. Consequently, the word 'data' is the most used unigram in the *description* texts of data engineers. Furthermore, since data engineers are part of a larger data science team, terms like 'project', 'team', 'development', 'support', and 'together' frequently appear in the *introduction* and *description* texts. As data

is the driver of digitalization, we can also find the terms 'big data' and 'analytics'. Moreover, the terms 'cloud' and 'analytics' complete the work of a data engineer building a bridge between both technology and data scientists. *Profile* descriptions asking for experience with data and architecture, cloud, and environments undermine this finding. Last, SQL, the most common language for data querying, is among the most used unigrams for data engineer vacancies.

Data Scientist: Data scientists have a more analytical focus and are responsible for model development and improvement. While the unigrams in the *introduction* texts are rather broad (i.e. ‘us’, ‘team’, ‘data’, ‘company’, ‘service’, ‘technology’), the *description* texts are more specific. Here, the texts focus on 'development', 'implementation', 'data', 'analytic', 'algorithm', and 'model'. Additionally, the majority of trigrams include either artificial intelligence or machine learning. Again, companies are looking for professional profiles focusing on data and data science, machine learning, statistics, and computer science. Besides the focus on SQL, it is often mandatory for data scientists to master Python or R.

Supplemental AI knowledge: The last vacancy type includes jobs that do not focus on ML or data science but mention that having experience with AI might be beneficial. These jobs might contain a small number of AI-related tasks but have a different focus, e.g., web developers or business consultants, who are among the ten most frequent unigrams for LEs. Thus, the *introduction* texts are more general and focus on technology, development, work, team, customer, or management. Additionally, these jobs might be more strategic than the more general jobs of data engineers or data scientists, and they often include phrases like 'mover', 'shaper' or '(shape) futur[e]'. Both languages, German and English, are among the most frequent unigrams in the *profile* texts for LEs, and knowledge, customer, and communication are also very frequent. On the contrary, SMEs mainly look for software developers or managers with a technical background.

5 Discussion

When screening all AI-related vacancies in our dataset, we found that digitization and digitalization are ongoing in Germany. As a result, there is a huge need for software developers and consultants to establish and optimize digital processes. Notably, we found an existing need for AI-based jobs. We identified six different types of AI-based jobs by utilizing the definition of [WDM17] for trainers, explainers, and sustainers and generalizing three different AI-based job types, i.e., data processor, data scientist, and jobs with supplemental AI knowledge. By manually labeling all relevant 935 vacancies, we found that the most significant share of vacancies does not require a deep understanding of AI (i.e., the jobs belonging to the supplemental AI knowledge). Instead, these vacancies only need an idea of what AI is and how to utilize it. Some vacancies even use the term AI as a buzzword together with Blockchain to express the need for adaptability to new technologies.

We could also see that the German vacancies incorporate the three future jobs predicted by [WDM17]. However, these vacancies account only for a small share of around 8% of all AI-based jobs. Furthermore, when screening all vacancies for labeling, we found that most job descriptions for trainers, explainers, and sustainers require a research background or are part of a research team. This fact hints that these job types are not yet part of the everyday data science team, and it might take some more time until a wider adoption of these jobs occurs.

When separating these job types between vacancies from SMEs and LEs we found that the share of trainers (4%-5%) and explainers (1%-2%) among all vacancies from the respective job type are almost equal for SMEs and LEs. On the contrary, we did not label the role of sustainers in any SME vacancy. This finding might indicate that SMEs might lag behind LEs when it comes to this more advanced job type. Another explanation could be that sustainers harmonize AI usage with governance, which might be less critical for SMEs due to lower organizational complexity than LEs. However, we need to mention that only 11% of the considered job offerings belong to SMEs, so there might be too few vacancies available in our dataset to draw a clear conclusion.

While we found almost no differences between the vacancies for a data processor between SMEs and LEs, we could see that the role of data scientists is more common for LEs than for SMEs. This finding indicates that the need for pure AI systems might be smaller for SMEs than for LEs in absolute and relative numbers. While more than half of the German employees work for SMEs [St22], the job role of data scientists is underrepresented. While this makes total sense since SMEs are often less complex than LEs, it hints that AI also plays a role for an average LE, while AI is primarily interesting for SMEs specializing in AI. This view is also consistent with the fact that the requirements for data scientists in SMEs are often more detailed and precise than in vacancies from LEs. LEs, on the contrary, do not look for pure data scientists only but for all-rounders.

Comparing the different distributions of job types between SMEs and LEs, we only found a significant difference between the vacancies looking for supplemental AI knowledge and data scientists utilizing a Mann-Whitney-U-Test ($p = .08$). For all other job types, we found no significant difference in the distribution between SMEs and LEs ($p = .3$). This finding follows the first indication when looking at Fig. 1.

When analyzing the need for knowledge of different programming languages, we found that Python is the primary programming language used in AI-based vacancies for LEs. This finding follows the trend that various AI research bases on Python, so it takes less adoption time for companies when new algorithms become available. Python is also a common requirement in SME vacancies, but they also ask for Java and C/C++. This more comprehensive variety of programming languages might indicate a fixed software stack in SMEs, i.e., some concentrate on Java, some on C, and others on Python. Due to the lower number of employees in SMEs, there might not always be a Python expert available in every company, so it might be easier to stick with the current software stack.

6 Conclusion

The role of AI is steadily gaining importance in the job market. By analyzing AI-related vacancies on one of the largest online vacancy platforms in Germany, we identify current trends in various AI-related jobs and specify differences between SMEs and large enterprises. We analyzed the spread of future AI jobs identified by a study by Accenture [WDM17] and found that there already is a need for the job roles of trainers, explainers, and sustainers in the German job market. While most vacancies originate from LEs, we found similar trends for SMEs, too. For instance, trainers and explainers make up 6% of all vacancies for the respective type of company.

Besides various similarities between SMEs and LEs, we also identified noticeable differences. For example, SMEs look for more specific profiles. They ask for a broader set of knowledge for different programming languages, while most LEs only ask for knowledge in Python.

A limitation of our work is that we only use a few keywords for our vacancy search on a single vacancy platform. Thus, we cannot guarantee that we found all AI-related vacancies. Especially job advertisements that include specific AI-related frameworks like TensorFlow but do not state the term AI in their descriptions are not part of our analysis. Additionally, relying only on a single data source implies that we do not cover vacancies that companies only list on different platforms. We acknowledge this limitation due to the necessary manual work to identify duplicates. Additionally, further platforms have different content classifications. For instance, it might not be possible to extract the job profile directly. However, due to our large sample of around 1,600 vacancies, we are confident that we incorporated the most important trends in our data. Additionally, we only analyzed the vacancies from a single point in time. A longer horizon might strengthen our results and allow us to identify changes in various trends.

Consequently, future research can redo our analysis later in time. Furthermore, future research might incorporate different data sources since we only use data from a single platform and vacancies from a single country, i.e., Germany. Additionally, future research can rely on alternative methodologies, e.g., conducting topic modeling via LDA.

Our results might help various stakeholders. First, it helps HR departments of SMEs and LEs of giving an overview of the requirements of different companies and competitors. This knowledge supports identifying new trends in AI-based vacancies and concising job descriptions. Second, the results might help educational institutions like training companies and universities (especially universities of applied sciences) to align syllabi with current requirements. A long-term study during different points in time might additionally help identifying new trends allowing earlier adoption and preparation. Last, the results also help applicants to get a broader picture of all hundreds of available vacancies. Thus, applicants can better filter vacancies according to their preferences.

7 Bibliography

[ALM03] Autor, D.H., Levy, F., Murnane, R.J.: The Skill Content of Recent Technological Change: An Empirical Exploration*. The Quarterly Journal of Economics 118, pp. 1279–1333, 2003.

[BD18] Bruun, E.P.G., Duka, A.: Artificial Intelligence, Jobs and the Future of Work: Racing with the Machines. Basic Income Studies 13, 2018.

[Be18] Bessen, J.: Artificial Intelligence and Jobs: The Role of Demand. In: The Economics of Artificial Intelligence: An Agenda. pp.291–307. University of Chicago Press, 2018.

[Bo03] Bostrom, N.: Ethical issues in advanced artificial intelligence. Science fiction and philosophy: from time travel to superintelligence 277, pp. 284, 2003.

[Bo18] Borgesius, Z.: Discrimination, artificial intelligence, and algorithmic decision-making. Strasbourg: Council of Europe, Directorate General of Democracy pp. 52, 2018.

[Bu20] Bundesregierung Deuschland: Fortschreibung der KI-Strategie, https://www.bundesregierung.de/breg-de/suche/fortschreibung-ki-strategie-1824340, last accessed 2022/10/05.

[Eu03] European Union: Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises, http://data.europa.eu/eli/reco/2003/361/oj, last accessed 2022/05/15.

[Ki18] Kim, M., Zimmermann, T., DeLine, R., Begel, A.: Data Scientists in Software Teams: State of the Art and Challenges. IIEEE Trans. Software Eng. 44, pp. 1024–1038, 2018.

[St22] Statistisches Bundesamt: Kleine und mittlere Unternehmen, https://www.destatis.de/DE/Themen/Branchen-Unternehmen/Unternehmen/Kleine-Unternehmen-Mittlere-Unternehmen/_inhalt.html, last accessed 2022/05/15.

[WDM17] Wilson, H.J., Daugherty, P.R., Morini-Bianzino, N.: The Jobs That Artificial Intelligence Will Create. MITSloan Management Review Summer 2017 Issue, pp. 5, 2017.

Anwendung von Machine Learning bei der datengetriebenen Prozessanalyse

Eine State-of-the-Art Literaturanalyse

Laslo Welz[1] und Helmut Beckmann[2]

Abstract: Die Disziplin Process Mining im Anwendungsbereich der datengetriebenen Prozessanalyse ermöglich die Abbildung realer Geschäftsprozesse durch die Extrahierung von Daten aus Eventlogs von Informationssystemen. Konventionelles Process Mining kann mit Ansätzen aus dem Bereich Machine Learning ergänzt werden, um die Prozessanalysen zu verbessern. Anhand einer Literaturanalyse untersucht diese Forschungsarbeit die Anwendung von Machine Learning bei Process Mining. Die Ergebnisse aus einer Stichprobe von 34 Publikationen zeigen, dass in den beiden Process Mining Bereichen „Discovery" und „Enhancement" die meisten Machine Learning-Methoden angewendet werden. Insbesondere ist die Anwendung von Entscheidungsbäumen und Neuronalen Netzen weit verbreitet.

Keywords: Datengetriebene Prozessanalyse, Process Mining, Machine Learning

1 Einleitung

Bei der datengetriebene Prozessanalyse, besser bekannt unter dem Begriff „Process Mining", werden verschiedene Disziplinen aus den Bereichen Prozessmodellierung, Prozessanalyse und Data Mining verknüpft. Das Ziel ist die Analyse von Prozessen anhand der Extraktion von Daten aus Eventlogs verschiedener Informationssysteme, welche ein reales Prozessmodell abbilden können. Darauf folgt in den meisten Fällen die Prozessangleichung und -optimierung. In den letzten Jahren ist Process Mining immer mehr in den Vordergrund gerückt, da Unternehmen Bedarf haben, ihre internen Prozesse im Detail zu analysieren. Auch kleine und mittlere Unternehmen (KMU) können durch Process Mining profitieren, sind aber bei der Anwendung durch meist unzureichender IT-Infrastruktur und qualitativ mangelhafter Eventlogs eingeschränkt [St21]. In der wissenschaftlichen Literatur werden verschiedene Felder und Anwendungen von Process Mining betrachtet [Ga19], wie beispielsweise in der Produktion [Du21], in der Logistik [BI17], oder im Bereich der Softwareentwicklung [KV17]. Es ist möglich, Process Mining anhand weiterer Analysemethoden aus dem Feld der Informatik zu kombinieren und zu verbessern

[1] Hochschule Heilbronn, Fakultät Wirtschaft und Verkehr, Max-Planck-Straße 39, 74081 Heilbronn, lawelz@stud.hs-heilbronn.de
[2] Hochschule Heilbronn, Fakultät Wirtschaft und Verkehr, Max-Planck-Straße 39, 74081 Heilbronn, helmut.beckmann@hs-heilbronn.de

doi:10.18420/inf2022_24

[ZSA21]. Ein sehr bekanntes Feld ist Machine Learning (ML). Das „Process Mining Manifesto“ nennt die Kombination von Process Mining mit anderen Analysemethoden als eine zentrale Herausforderung im Forschungsbereich Process Mining [va12]. Auch vom Brocke et al. [vo21] sehen Forschungsbedarf bei der Ergänzung von Process Mining mit weiteren technologischen Ansätzen und deren erfolgreichen Adaption in Unternehmen. Gerade im Bereich ML benötigen insbesondere KMUs das nötige Know-how für eine erfolgreiche Umsetzung fortgeschrittener Analysemethoden [BvK20]. Die Bereiche ML und Process Mining eröffnen somit ein weiteres Forschungsspektrum, welches in der Literatur bereits einige Ansätze aufzeigt [EYR16], aber nicht grundlegend auf die möglichen Anwendungen von ML in allen Bereichen vom Process Mining abzielt. Aus der Problemstellung ergibt sich folgende Fragestellung: *Welche Anwendungen für Machine Learning bei der datengesteuerten Prozessanalyse können identifiziert werden?*

Ziel dieser Forschungsarbeit ist es mithilfe einer strukturierten Literaturanalyse den Status Quo bei Anwendungen von Machine Learning in Verbindung mit Process Mining zu identifizieren. Die Ergebnisse werden erst quantitativ und im Anschluss für jeden Process Mining Bereich im Detail dargestellt. Nachfolgend zur Problemstellung und Forschungsfrage wird in Kapitel 2 ein umfangreiches Verständnis über die Bereiche Process Mining und Machine Learning erarbeitet. Kapitel 3 beschäftigt sich mit der anzuwendenden Methodik für die Literaturanalyse. Im nächsten Schritt werden in Kapitel 4 anhand der identifizierten Literatur, Kategorien von ML-Anwendungen innerhalb der drei Bereiche Process Discovery, Process Conformance, sowie Process Enhancement bestimmt und mithilfe einer Konzeptmatrix strukturiert. Im Anschluss werden die Ergebnisse der Literaturanalyse in Kapitel 5 zusammengefasst und diskutiert, sowie ein Ausblick auf zukünftige Forschungsansätze gegeben.

2 Grundlagen und Begriffe

Im folgenden Kapitel werden die beiden Begriffe „Process Mining“ und „Machine Learning“ näher dargestellt und erläutert, um die Grundlage für diese Forschungsarbeit zu etablieren.

2.1 Process Mining

Process Mining ist im Vergleich zum Data Mining nicht Daten- sondern Prozess-orientiert. Das Ziel ist die Analyse von Prozessen anhand der Extraktion von Eventlogs verschiedener Informationssysteme und der anschließenden Prozessangleichung und -optimierung [va11]. Ein Event steht immer in einer Beziehung zu einer Prozessaktivität, welche beispielsweise einen bestimmten Arbeitsschritt darstellt. Ebenso enthält ein Event mehrere Attribute, welche zumindest aus einer ID, einem Namen und einem Zeitstempel bestehen. Lassen sich mehrere Events zeitlich hintereinander zu einer bestimmten Prozessinstanz zuordnen, so spricht man von einem Case oder einem „Trace of Events“ [KJS21].

Drei Bereiche von Process Mining können identifiziert werden. Process Discovery ermöglicht die Abbildung von Ist-Prozessen durch die Verarbeitung von Event-Logs aus beteiligten Informationssystemen. So kann ein Prozessmodell, unabhängig von bereits bestehenden Informationen, erstellt werden. Bei der Process Conformance werden der nachgebildete Prozess mit dem eigentlichen Soll-Prozess verglichen. Durch den Abgleich kann geprüft werden, inwieweit das Prozessmodell definierte Regelungen einhält, oder in welcher Form Abweichungen, an welcher Prozesssequenz, in welchem Maße auftreten. Zuletzt dient die Process Enhancement dazu, das Prozessmodell anhand des abgeleiteten Wissens zu verbessern oder zu erweitern. Bei einer Verbesserung können beispielsweise zwei redundante Prozessabfolgen, zu einer einzelnen Prozessabfolge kombiniert werden. Eine Erweiterung kann das bestehende Prozessmodel durch die Zugabe bestimmter Messkriterien evaluiert werden, oder durch ML-Anwendungen Vorhersagen zu fortlaufenden Prozessschritten gemacht werden [va11].

2.2 Machine Learning

ML ist ein Teilbereich der Künstlichen Intelligenz, welche wiederum ein Teilbereich der Informatik ist. Einige Teile von ML werden aber auch dem Statistik- und Mathematikbereich zugeordnet. ML beschäftigt sich mit dem Erstellen und Trainieren von mathematischen Modellen, anhand von Beispieldaten, um bestimmte Aufgaben zu lösen, ohne diese explizit dafür zu programmieren [Zh20]. Laut Murphy [Mu12] besteht ML aus einer Vielzahl von Methoden, die aus historischen Daten automatisch Muster ableiten und diese wiederum nutzen, um Vorhersagen auf neue Daten zu initiieren, sowie bei der Entscheidungsfindung zu unterstützen. Innerhalb von ML existieren verschiedene Lernmethoden, zwei davon sind das Überwachte Lernen und das unüberwachte Lernen. Beim überwachten Lernverfahren lernt das Modell bestimmte Regeln anhand gelabelter Daten, die den gewünschten Output vorgeben. Nach dem Training wird versucht, diese Regeln auf neue Daten ohne Label anzuwenden. Durch die Anwendung von Performance Metriken, kann das Model evaluiert werden. Beispiele für überwachte Lernmethoden sind Klassifikations- oder Regressionsalgorithmen. Anders als bei der überwachten Methode werden beim unüberwachten Lernen die Daten im Vorhinein ohne Labels herangenommen. Dadurch, dass kein Output gegeben ist, wird versucht die Daten anhand von ähnlichen Mustern und Strukturen zusammenzuführen und die darunterliegenden Einflussfaktoren zu identifizieren. Für einen unüberwachten Lösungsansatz können Clustermethoden wie beispielsweise K-Means verwendet werden [Zh20]. Ein Teilbereich von ML ist Deep Learning. Dieser beschäftigt sich mit der Anwendung komplexerer Algorithmen, welche insbesondere bei der Verarbeitung von Text-, Bild- oder Sprachdaten genutzt werden. Ein Modell kann aus mehreren Schichten, Ebenen und Konnektoren bestehen, wie beispielsweise Neuronale Netze, welche die Funktionsweise eines menschlichen Gehirns imitieren [Zh20].

3 Methodik zur Literaturanalyse

Zur Eingrenzung der zu betrachteten Literatur wird die strukturierte und nachvollziehbare Suchmethodik nach vom Brocke et al. angewendet [vo09]. Diese beinhaltet die Journal-, Datenbank- und Keywordsuche [vo09]. Ergänzend wird die Vorwärts- und Rückwärtssuche nach Webster und Watson angewendet [WW02]. Die Erfassung der Anwendungskategorien ist durch die Erstellung einer Konzeptmatrix zu realisieren. Dazu ist die systematische Methodik nach Webster und Watson durchzuführen [WW02]. Für die Literatursuche innerhalb dieser Forschungsarbeit besteht der Fokus auf Literaturdatenbanken im Bereich der Wirtschaftsinformatik und Informatik. Insgesamt werden die vier zugänglichen Datenbanken *ACM, IEEE Xplore, SpringerLink* und *ScienceDirect* betrachtet. Zusätzlich dazu wird nach Veröffentlichungen innerhalb zwei internationaler Konferenzen im Bereich Prozessmanagement recherchiert. Die International Conference on Business Process Management (ICBPM) enthält in den Jahren 2017-2021 eigene Tracks für Process Mining und Process Discovery. Seit 2019 existiert auch die International Conference on Process Mining (ICPM). Diese enthält seit 2020 einen eigenen Workshop zum Einsatz von ML in Process Mining (ML4PM), welcher für das Jahr 2020 untersucht wird. Die zuvor vorgestellten Literaturdatenbanken und Konferenzen werden mit einem passenden Suchterm, entsprechend dem Thema dieser Forschungsarbeit, durchsucht. Der folgende Suchterm wird angewendet: *(“Process Mining” OR “Process Discovery” OR “Process Intelligence”) AND “Machine Learning”.*

Die Suche wurde im Dezember 2021 durchgeführt und ergab 857 Treffer für den Suchzeitraum von 2016-2021. Von diesen Treffern wurden 213 Publikationen genauer durch ein Screening des Titels und Abstracts betrachtet. 34 thematisch relevante Publikationen konnten auf der Grundlage der festgelegten Ein- und Ausschlusskriterien ausgewählt werden, wie beispielsweise die explizite Anwendung von ML-Methoden in Process Mining. Alle 34 Publikationen wurden für die Einordung in die Konzeptmatrix einem Peer-Review unterzogen.

4 Ergebnisse der Literaturanalyse

Durch die Konzeptmatrix können die in den Publikationen identifizierten ML-Anwendungen systematisch dargestellt werden. Zuerst werden die ML-Anwendungen den jeweiligen Process Mining Bereichen „Discovery“, „Conformance“ und „Enhancement“ zugewiesen. Hierzu wird ebenfalls vermerkt, ob der entsprechende Anwendungsbereich explizit in der Publikation erwähnt wird. Für jeden Bereich werden die angewendeten Methoden in die Kategorie „Überwacht“ oder „Unüberwacht“ unterteilt. Zusätzlich werden Anwendungen aus dem ML-Teilbereich „Deep Learning“ als eigenständige Kategorie erfasst. Ebenso wird anhand der Profession der Autoren und der Datenherkunft bestimmt, ob die Publikation einem akademischen oder praktischen Ursprung entspricht. Die Zuordnung der ML-Anwendungen mit den entsprechenden Publikationen erfolgt in der Kon-

zeptmatrix (siehe Abb. 1). Im folgenden Abschnitt werden allgemeine quantitative Ergebnisse, als auch die identifizierten ML-Anwendungen für jeden Process Mining Bereich im Detail, dargestellt.

		Ursprung			Discovery			Conformance			Enhancement		
#	Publikation	Akademisch	Praktisch	Explizite Erwähnung des Bereichs?	Überwacht	Unüberwacht	Deep Learning	Überwacht	Unüberwacht	Deep Learning	Überwacht	Unüberwacht	Deep Learning
#1	[LW18]	X		N							X	X	
#2	[MAM19]		X	N							X		
#3	[Ne21]	X		J					X				
#4	[TCK20]	X		N							X		
#5	[Ve17]		X	J				X			X	X	
#6	[EK21]	X		J	X								
#7	[EYR16]	X		J							X		
#8	[FP19]	X		N		X	X					X	X
#9	[KJS21]	X		J									X
#10	[SNM18]	X		J		X							
#11	[SZS18]	X		J			X						
#12	[TJ21]	X		J					X				
#13	[Ta20]	X		N									X
#14	[Te19]	X	X*	N	X	X		X	X	X			
#15	[We21]		X	J							X		
#16	[NSM18]	X		J			X			X			
#17	[Se19]	X		N								X	
#18	[CDG19]	X		N			X						X
#19	[SMF21]	X		J	X		X						
#20	[AM16]	X		J		X							
#21	[HKG20]	X		J			X			X			X
#22	[Ju20]	X		J				X					
#23	[Zh19]	X		J	X	X							
#24	[BBC18]	X	X*	J		X							
#25	[Su20]	X	X*	N							X		
#26	[Te18]	X		J			X						
#27	[BRK20]	X		J	X								
#28	[BAD21]	X		N									X
#29	[Ng21]	X		N									X
#30	[Ch21]	X		N									X
#31	[BCH21]	X		J				X		X			
#32	[QCP21]	X		N	X								
#33	[Lu21]	X		J								X	
#34	[PvW21]	X		J						X			
∑ (absolut)					6	6	6	4	3	5	7	5	8
∑ (relativ)					18%	18%	18%	12%	9%	15%	21%	15%	24%

*Publikation ist akademisch und verwendet Daten eines Unternehmens
J=Ja / N=Nein

Abb. 1: Konzeptmatrix (eigene Darstellung)

4.1 Allgemein

Die Betrachtung hinsichtlich des Ursprungs zeigt eine starke akademische Ausprägung. Nur drei der 34 betrachteten Publikationen weisen einen praktischen Bezug auf. Weitere drei Publikationen machen dies nicht direkt, aber verwenden für ihren Forschungsansatz Daten eines Unternehmens. Insgesamt werden 42 ML-Anwendungen im Kontext Process Mining erwähnt. Insbesondere in den Bereichen „Enhancement“ (20-mal) und „Discovery“ (18-mal) werden vermehrt ML-Methoden angewendet. Im Bereich „Conformance“ können mit zwölf Anwendungen die wenigsten ML-Methoden identifiziert werden. Etwas weniger als zwei Drittel der Publikationen (61,7%) erwähnen den Anwendungsbereich explizit. Die drei ML-Kategorien sind fast gleich stark vertreten. Es werden 19 der Kategorie „Deep Learning“, 17 ML-Anwendungen der Kategorie „Überwacht“ und 14 der unüberwachten Lernmethodik zugeordnet. Insgesamt werden 19 verschiedene ML-Methoden angewendet. Dabei können gleiche Methoden für mehrere Bereiche angewendet werden und pro Anwendung mehrere ML-Methoden genutzt werden können. In der Kategorie „Überwacht“ wurden mit acht ML-Methoden die meisten erwähnt. Die dort meistgenutzte Methode ist der Entscheidungsbaum (fünfmal). Bei Unüberwachten ist es K-Means-Clustering (viermal) und bei Deep Learning wird ein Long Short-Term Memory (LSTM)-Modell am häufigsten angewendet (achtmal).

4.2 Process Discovery

Bei der Process Discovery werden ML-Methoden angewendet, um die datengetriebene Prozessmodellierung aus Eventlogs zu unterstützen, oder zu verbessern. Die Autoren in [SNM18] und [AM16] nutzen sogenannte Trace Clustering Algorithmen, welche Eventlogs in kleinere Instanzen unterteilen, um so die Komplexität des Modellierungsalgorithmus zu reduzieren. Ähnliche Ansätze nutzen hierzu Deep Learning Methoden, wie Recurrent Neural Networks (RNN) [SZS18] und LSTM-Modelle [Te18]. Ein weiterer Problemansatz, für welcher ML genutzt werden kann, ist die Verarbeitung von wenig qualitativen Eventlogs, für die Erstellung eines Prozessmodells [FP19]. Bestehende Process Discovery Algorithmen werden meist mit einem unüberwachten Lernverfahren erstellt. Im Vergleich dazu schlagen Sommers et al. [SMF21] vor, die Prozessmodellierung anhand eines überwachten Modell-Trainings mit Graph Convolutional Neural Networks (GCN) durchzuführen, um deutlich realitätsgetreuere Prozesse abbilden zu können. Ebenso kann mit überwachten Lernmethoden untersucht werden, inwieweit Prozessmodelle durch laufende Veränderungen beeinflusst werden [BRK20]. Auch bei Problemen mit spezifischen Anwendungsfällen lassen sich ML-Methoden als Lösungsansatz anwenden. Das kann sowohl die Anwendung von Process Discovery auf spezielle Informationssysteme [Zh19] als auch die Extraktion von unstrukturierten Datenformaten, wie ganzen Texten, sein [BBC18; QCP21]. Ein weiterer Spezialfall für die Anwendung von ML ist die Bestimmung von einzelnen Prozessaktivitäten, die mehrere Eventlogs referenzieren [Te19]. Für nachfolgende Process Mining Operationen, wie beispielsweise Process Enhancement, können ML-Methoden bereits bei der Discovery benötigte Merkmale in Daten ausfindig machen [EK21].

4.3 Process Conformance

Innerhalb der Process Conformance nutzen die zugeordneten Publikationen verschiedene ML-Methoden für die Anomalieerkennung in bestehenden Prozessmodellen. Diese betrachten unter anderem das Problem, dass der Abgleich eines aus Eventlogs erstellen Prozessmodells mit einem definierten Soll-Prozess nur mithilfe eines domänenspezifischen Vorwissens möglich ist. Neto et al. [Ne21] nutzen eine Online-Clustermethode, die Anomalieerkennung ohne spezifisches Vorwissen, auf ein Prozessmodell anwendet. Im Gegenteil dazu erwägt Junior et al. [Ju20] einen Klassifikationsansatz und ergänzt diesem um eine Natural Language Processing (NLP)-Komponente, um Zusammenhänge zwischen verschiedenen Eventlogs besser darzustellen. Ebenso kann ein Entscheidungsbaum-Modell angewendet werden, um die Performance der Anomalieerkennung deutlich zu verbessern [TJ21]. Eine weitere Anwendungsmöglichkeit bilden Deep Learning-Methoden für die Anomalieerkennung auf der Ebene einzelner Aktivitäten. Hierbei werden verschiedene Modelle angewendet, wie Generative Adversarial Nets (GAN) [Ta20], RNN [NSM18; PvW21] und LSTM [BCH21; HKG20]. Ein wichtiger Aspekt der Process Conformance ist die eindeutige Erkennung von Prozessverstößen. Die Autoren in [Ve17] entwickeln auf diesem Aspekt ein ML-System, welches Prozessverstöße erkennt und zu diesen eine Handlungsempfehlung ausgeben.

4.4 Process Enhancement

Der letzte Bereich, Process Enhancement, beschäftigt sich mit der Verbesserung oder Erweiterung des extrahierten Prozessmodells. ML-Methoden werden hier angewendet, um potenzielle nächste Prozessschritte vorherzusagen. Für die Vorhersage können sowohl überwachte als auch unüberwachte Lernmethoden angewendet werden [LW18]. Die Autoren in [TCK20] untersuchen hierfür verschiedene Klassifikationsmethoden und empfehlen in ihrer Forschung die Anwendung von Entscheidungsbaummodellen. Ebenso wird im Bereich Deep Learning vermehrt ein LSTM-Modell für die Vorhersage nächster Prozessausführungen genutzt [CDG19; FP19; HKG20; Ng21; Ch21]. Im Speziellen betrachten Käppel et al. [KJS21] die Vorhersage mit kleinen Datensätzen. Die Autoren in [BAD21] nutzen ein RNN, um Vorhersagen aus unvollständigen Event Traces zu erhalten. Des Weiteren kann durch die Anwendung eines Naive Bayes Klassifikators vorhergesagt werden, wie lange die Durchlaufzeit des Prozesses sein wird, ohne dass dieser vollständig durchgelaufen ist [We21]. Ebenfalls ist es möglich, die verbleibende Zeit für parallellaufende Prozesse vorherzusagen, um Aktivitäten danach proirisieren zu können. Sun et al. [Su20] nutzen hierfür ein Modell mit einem Extreme Gradient Boosting (XGBoost) Klassifikator. Für Managementzwecke können aus Prozessanalysen wichtige operative oder strategische Entscheidungen abgeleitet werden. Durch ML-Anwendungen können solche Entscheider-Systeme datenbasiert Vorhersagen generieren [MAM19]. Dazu werden von den Autoren in [EYR16] verschiedene Modelle, basierend auf Entscheidungsbäumen angewendet. Neben harten Fakten können automatisch Empfehlungen an den Anwender einer ML-gesteuerten Prozessanalyse weitergegeben werden [Ve17]. Seeliger et al. [Se19] stellen

eine interaktive Anwendungssoftware, welche Prozessanalysen in Echtzeit darstellt und anhand eines Clustermodells Empfehlungen ausgibt.

5 Diskussion und Ausblick

Mithilfe der Literaturanalyse nach vom Brocke et al. und der Einordung der Literatur in eine Konzeptmatrix nach Webster und Watson konnten ML-Anwendungen bei Process Mining entlang der drei Bereiche identifiziert werden. Die Zuordnung der jeweiligen ML-Anwendung in die Konzeptmatrix kann durch die mehrheitlich explizite Erwähnung des einzugrenzenden Process Mining Anwendungsbereichs validiert werden. Für die Beantwortung der Forschungsfrage lässt sich feststellen, dass die Anwendung von ML-Methoden der Kategorie „Deep Learning“ am häufigsten erfolgt. Die vermehrte Nutzung von Deep Learning-Methoden kann durch die bessere Modellperformance und Skalierbarkeit sowie die einfachere Übertrag- und Anpassbarkeit gegenüber klassischer ML-Methoden begründet werden [Au19]. Es lassen sich für die Bereiche „Discovery“ und „Enhancement“ mehrere Anwendungsfälle für den Einsatz von ML-Methoden identifizieren. Bei der Process Discovery können Prozessmodelle durch die Anwendung ML-Methoden deutlich effizienter und realitätsnäher dargestellt werden. Speziell für KMUs können ML-Methoden zur Verbesserung der Eventlogqualität [FP19] von Vorteil sein, um fehlende Skalenvorteile auszugleichen. Die Vorhersage nächster Prozessschritte in der Enhancement Phase findet Anwendung bei einer Vielzahl von ML-Methoden. Einer dieser kann nach dem Ansatz von Käppel et al. [KJS21] bei KMUs genutzt werden, die nur kleine Datensätze zur Verfügung haben. Für die Process Conformance werden verschiedene ML-Anwendungen hauptsächlich für die Anomalieerkennung genutzt. Obwohl Process Mining eine wachsende praktische Disziplin in Unternehmen darstellt, sind Publikationen zum Thema ML in Process Mining eher akademischen Ursprungs. Damit stellt sich die Frage, inwieweit die untersuchten Methoden für die direkte Anwendung in Unternehmen, insbesondere für KMUs, geeignet sind. Eine Untersuchung der Anwendbarkeit dieser ML-Methoden bei Process Mining Projekten und möglicher Herausforderungen bei der Implementierung im Unternehmen wäre sinnvoll. Eine ähnliche Forschungslücke wird auch von den Autoren in [vo21] erwähnt. Des Weiteren kann mit der aktuell kleinen Stichprobe keine quantitative Aussage zur Anwendung von ML-Methoden in Process Mining getroffen werden, weshalb diese erweitert werden sollte. Mit der aktuellen Relevanz des Themas in der Forschung, insbesondere durch eigene Konferenztracks und Workshops, können auf Basis dieser Arbeit in Zukunft weitere Ansätze ergänzt werden. Ebenfalls interessant wäre eine detaillierte Betrachtung eines einzelnen Process Mining Teilbereichs, beispielsweise bei der Process Discovery oder der Process Enhancement. Auch eine explizite Untersuchung des Bereichs Process Conformance wäre möglich, um weitere ML-Anwendungsfälle für diesen Bereich zu identifizieren. Weitere Forschung zur Anwendung von Deep Learning Methoden bei Process Mining kann weitere Erkenntnisse für Gründe und Probleme solcher Ansätze hervorbringen.

6 Literaturverzeichnis

[AM16] Appice, A.; Malerba, D.: A Co-Training Strategy for Multiple View Clustering in Process Mining. IEEE Transactions on Services Computing 6/9, S. 832–845, 2016.

[Au19] Augusto, A. et al.: Automated Discovery of Process Models from Event Logs: Review and Benchmark. IEEE Transactions on Knowledge and Data Engineering 4/31, S. 686–705, 2019

[BAD21] Baskharon, F.; Awad, A.; Di Francescomarino, C.: Predicting Remaining Cycle Time from Ongoing Cases: A Survival Analysis-Based Approach. In (Leemans, S.; Leopold, H. Hrsg.): Process Mining Workshops. Springer International Publishing, Cham, S. 99–111, 2021.

[BBC18] Banziger, R. B.; Basukoski, A.; Chaussalet, T.: Discovering Business Processes in CRM Systems by Leveraging Unstructured Text Data: 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, S. 1571–1577, 2018.

[BCH21] Boltenhagen, M.; Chetioui, B.; Huber, L.: An Alignment Cost-Based Classification of Log Traces Using Machine-Learning. In (Leemans, S.; Leopold, H. Hrsg.): Process Mining Workshops. Springer International Publishing, Cham, S. 136–148, 2021.

[BI17] Becker, T.; Intoyoad, W.: Context Aware Process Mining in Logistics. Procedia CIRP 63, S. 557–562, 2017.

[BRK20] Baier, L.; Reimold, J.; Kuhl, N.: Handling Concept Drift for Predictions in Business Process Mining: 2020 IEEE 22nd Conference on Business Informatics (CBI). IEEE, S. 76–83, 2020.

[BvK20] Bauer, M.; van Dinther, C.; Kiefer, D.: Machine Learning in SME: An Empirical Study on Enablers and Success Factors. AMCIS 2020 Proceedings 3, 2020.

[CDG19] Camargo, M.; Dumas, M.; González-Rojas, O.: Learning Accurate LSTM Models of Business Processes. In (Hildebrandt, T. et al. Hrsg.): Business Process Management. Springer International Publishing, Cham, S. 286–302, 2019.

[Ch21] Chiorrini, A. et al.: A Preliminary Study on the Application of Reinforcement Learning for Predictive Process Monitoring. In (Leemans, S.; Leopold, H. Hrsg.): Process Mining Workshops. Springer International Publishing, Cham, S. 124–135, 2021.

[Du21] Dunzer, S. et al.: The Status Quo of Process Mining in the Industrial Sector. In (Ahlemann, F.; Schütte, R.; Stieglitz, S. Hrsg.): Innovation Through Information Systems. Springer International Publishing, Cham, S. 629–644, 2021.

[EK21] Elkhovskaya, L.; Kovalchuk, S.: Feature Engineering with Process Mining Technique for Patient State Predictions. In (Paszynski, M. et al. Hrsg.): Computational Science – ICCS 2021. Springer International Publishing, Cham, S. 584–592, 2021.

[EYR16] Es-Soufi, W.; Yahia, E.; Roucoules, L.: On the Use of Process Mining and Machine Learning to Support Decision Making in Systems Design. In (Harik, R. et al.

Hrsg.): Product Lifecycle Management for Digital Transformation of Industries. Springer International Publishing, Cham, S. 56–66, 2016.

[FP19] Folino, F.; Pontieri, L.: Pushing More AI Capabilities into Process Mining to Better Deal with Low-Quality Logs. In (Di Francescomarino, C.; Dijkman, R.; Zdun, U. Hrsg.): Business Process Management Workshops. Springer International Publishing, Cham, S. 5–11, 2019.

[Ga19] Garcia, C. d. S. et al.: Process mining techniques and applications – A systematic mapping study. Expert Systems with Applications 133, S. 260–295, 2019.

[HKG20] Hanga, K. M.; Kovalchuk, Y.; Gaber, M. M.: A Graph-Based Approach to Interpreting Recurrent Neural Networks in Process Mining. IEEE Access 8, S. 172923–172938, 2020.

[Ju20] Junior, S. B. et al.: Anomaly Detection on Event Logs with a Scarcity of Labels: 2020 2nd International Conference on Process Mining (ICPM). IEEE, S. 161–168, 2020.

[KJS21] Käppel, M.; Jablonski, S.; Schönig, S.: Evaluating Predictive Business Process Monitoring Approaches on Small Event Logs. In (Paiva, A. C. R. et al. Hrsg.): Quality of Information and Communications Technology. Springer International Publishing, Cham, S. 167–182, 2021.

[KV17] Keith, B.; Vega, V.: Process mining applications in software engineering. In (Mejia, J. et al. Hrsg.): Trends and Applications in Software Engineering. Springer International Publishing, Cham, S. 47–56, 2017.

[Lu21] Luettgen, S. et al.: Case2vec: Advances in Representation Learning for Business Processes. In (Leemans, S.; Leopold, H. Hrsg.): Process Mining Workshops. Springer International Publishing, Cham, S. 162–174, 2021.

[LW18] Liu, Q.; Wu, B.: Prediction of Business Process Outcome based on Historical Log. In (Unknown Hrsg.): Proceedings of the 10th International Conference on Computer Modeling and Simulation - ICCMS 2018. ACM Press, New York, New York, USA, S. 118–122, 2018.

[MAM19] Mesabbah, M.; Abo-Hamad, W.; McKeever, S.: A hybrid process mining framework for automated simulation modelling for healthcare: 2019 Winter Simulation Conference, S. 1091–1102, 2019.

[Mu12] Murphy, K. P.: Machine learning. A probabilistic perspective. The MIT Press, Cambridge, Massachusetts, London, England, 2012.

[Ne21] Neto, R. V. et al.: On the use of online clustering for anomaly detection in trace streams. In (Araújo, R. D. et al. Hrsg.): XVII Brazilian Symposium on Information Systems. ACM, New York, NY, USA, S. 1–8, 2021.

[Ng21] Nguyen, A. et al.: Time Matters: Time-Aware LSTMs for Predictive Business Process Monitoring. In (Leemans, S.; Leopold, H. Hrsg.): Process Mining Workshops. Springer International Publishing, Cham, S. 112–123, 2021.

[NSM18] Nolle, T.; Seeliger, A.; Mühlhäuser, M.: BINet: Multivariate Business Process Anomaly Detection Using Deep Learning. In (Weske, M. et al. Hrsg.): Business Process Management. Springer International Publishing, Cham, S. 271–287, 2018.

[PvW21] Peeperkorn, J.; van den Broucke, S.; Weerdt, J. de: Supervised Conformance Checking Using Recurrent Neural Network Classifiers. In (Leemans, S.; Leopold, H. Hrsg.): Process Mining Workshops. Springer International Publishing, Cham, S. 175–187, 2021.

[QCP21] Quishpi, L.; Carmona, J.; Padró, L.: Improving the Extraction of Process Annotations from Text with Inter-sentence Analysis. In (Leemans, S.; Leopold, H. Hrsg.): Process Mining Workshops. Springer International Publishing, Cham, S. 149–161, 2021.

[Se19] Seeliger, A. et al.: ProcessExplorer: Intelligent Process Mining Guidance. In (Hildebrandt, T. et al. Hrsg.): Business Process Management. Springer International Publishing, Cham, S. 216–231, 2019.

[SMF21] Sommers, D.; Menkovski, V.; Fahland, D.: Process Discovery Using Graph Neural Networks: 2021 3rd International Conference on Process Mining (ICPM). IEEE, S. 40–47, 2021.

[SNM18] Seeliger, A.; Nolle, T.; Mühlhäuser, M.: Finding Structure in the Unstructured: Hybrid Feature Set Clustering for Process Discovery. In (Weske, M. et al. Hrsg.): Business Process Management. Springer International Publishing, Cham, S. 288–304, 2018.

[St21] Stertz, F. et al.: Expectations vs. Experiences – Process Mining in Small and Medium Sized Manufacturing Companies. In (Polyvyanyy, A. et al. Hrsg.): Business Process Management Forum. Springer International Publishing, Cham, S. 195–211, 2021

[Su20] Sun, X. et al.: Remaining Time Prediction of Business Processes based on Multilayer Machine Learning: 2020 IEEE International Conference on Web Services (ICWS). IEEE, S. 554–558, 2020.

[SZS18] Shunin, T.; Zubkova, N.; Shershakov, S.: Neural Approach to the Discovery Problem in Process Mining. In (van der Aalst, W. M. P. et al. Hrsg.): Analysis of Images, Social Networks and Texts. Springer International Publishing, Cham, S. 261–273, 2018.

[Ta20] Taymouri, F. et al.: Predictive Business Process Monitoring via Generative Adversarial Nets: The Case of Next Event Prediction. In (Fahland, D. et al. Hrsg.): Business Process Management. Springer International Publishing, Cham, S. 237–256, 2020.

[TCK20] Tama, B. A.; Comuzzi, M.; Ko, J.: An Empirical Investigation of Different Classifiers, Encoding, and Ensemble Schemes for Next Event Prediction Using Business Process Event Logs. ACM Transactions on Intelligent Systems and Technology 6/11, S. 1–34, 2020.

[Te18] Tello-Leal, E. et al.: Predicting Activities in Business Processes with LSTM Recurrent Neural Networks. 2018 ITU Kaleidoscope: Machine Learning for a 5G Future, S. 1–7, 2018.

[Te19] Tello, G. et al.: Machine Learning-Based Framework for Log-Lifting in Business Process Mining Applications. In (Hildebrandt, T. et al. Hrsg.): Business Process Management. Springer International Publishing, Cham, S. 232–249, 2019.

[TJ21] Tavares, G. M.; Junior, S. B.: Process Mining Encoding via Meta-learning for an Enhanced Anomaly Detection. In (Bellatreche, L. et al. Hrsg.): New Trends in Database and Information Systems. Springer International Publishing, Cham, S. 157–168, 2021.

[va11] van der Aalst, W. M. P.: Process Mining. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[va12] van der Aalst, W. et al.: Process Mining Manifesto. In (Daniel, F.; Barkaoui, K.; Dustdar, S. Hrsg.): Business Process Management Workshops. Springer Berlin Heidelberg, Berlin, Heidelberg, S. 169–194, 2012.

[Ve17] Veit, F. et al.: The Proactive Insights Engine: Process Mining meets Machine Learning and Artificial Intelligence: 15th International Conference on Business Process Management 2017, S. 1–5, 2017.

[vo09] vom Brocke, J. et al.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process: 17th European Conference on Information Systems (ECIS), 2009.

[vo21] vom Brocke, J. et al.: A Five-Level Framework for Research on Process Mining. Business & Information Systems Engineering 5/63, S. 483–490, 2021.

[We21] Welsing, M. et al.: Combining Process Mining and Machine Learning for Lead Time Prediction in High Variance Processes. In (Behrens, B.-A. et al. Hrsg.): Production at the leading edge of technology. Springer Berlin Heidelberg, Berlin, Heidelberg, S. 528–537, 2021.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 2/26, S. xiii–xxiii, 2002.

[Zh19] Zhu, R. et al.: Automatic Real-Time Mining Software Process Activities From SVN Logs Using a Naive Bayes Classifier. IEEE Access 7, S. 146403–146415, 2019.

[Zh20] Zhang, X.-D.: Machine Learning. In (Zhang, X.-D. Hrsg.): A Matrix Algebra Approach to Artificial Intelligence. Springer Singapore; Imprint Springer, Singapore, S. 223–440, 2020.

[ZSA21] Zerbino, P.; Stefanini, A.; Aloini, D.: Process Science in Action: A Literature Review on Process Mining in Business Management. Technological Forecasting and Social Change 172, S. 121021, 2021.

A Demonstration System towards NLP and Knowledge Driven Data Platforms for Civil Engineering

Janos Borst[1], Christofer Meinecke[1], Daniel Wiegreffe[1], Andreas Niekler[1]
{firstname.lastname@uni-leipzig.de}

Abstract: We present a demonstrator that shows a concept towards a smart document platform for Civil Engineering. It demonstrates NLP-supported organisation of documents included into a graphical user interface. The demonstrator addresses the fundamental problem of data structure that includes the internal project logic of a planning project in Civil Engineering. To this end, we build a knowledge graph that includes not only domain-specific knowledge but also standardised project structures[2]. So far included functionalities are a navigation that allows filtering according to a logic familiar to engineers and a simplified data import via automatic classification and tagging using language technologies.

1 Introduction

Many work processes in Civil Engineering (CE) are increasingly standardized and can be supported by algorithms. Often, only a small part of a project as a whole still offers room for non-standardized solutions. In such a standardized workflow, a lot of documentation and planning documents are generated. This data must remain findable, usable and reusable in the work process. There is major potential for automating the processing of the large volumes of data that are generated in the course of various projects. In detail, this concerns the documentation of planning bases and results of expert activities, which are mainly provided in the form of unstructured documents. In the further course of the project work in CE, offers and invoices are generated, and the construction process is documented for the subsequent management of the constructed facilities. The current workflow requires that all companies involved in a planning project have the same document basis. However, each company operates its own database for this purpose. This creates redundancies, problems with data protection are foreseeable, and versioning and managing revisions is nearly impossible. In addition, some documents retain validity for other projects and must remain available across project contexts and extended time periods. These multiple dependencies create a stakeholder network for certain documents in the planning process.

A shared and uniform data platform would be an enormous relief for the entire industry sector since no solitary document storage solutions would be necessary. In a data trust, a platform assumes responsibility for data protection, versioning, findability and linking of

[1] Leipzig University, Faculty of Mathematics and Computer Science, Augustusplatz 10, 04109 Leipzig
[2] Graph available under https://doi.org/10.5281/zenodo.5055839 [Bo21]

doi:10.18420/inf2022_25

information, so that collaborative and efficient work is possible. In terms of digitization, concrete work facilitation would arise from unified ways of working abstracted from the physical drive, abstraction of engineers' work paradigms and easy access management. Since the work logic and paradigms can be predefined in such a platform, assistants could be assigned to the data work instead of expensive engineers.

In this work, we present a demonstrator[3] that shows a perspective towards a smart document platform for CE. We describe an implementation based on the very strict guidelines that are standardized in Germany. We contribute a data structure that addresses the internal project logic of a planning project, a navigation that allows filtering according to a logic familiar to engineers and a simplified data import via automatic classification and tagging.

2 Related Work

Our system is based on four main aspects: Knowledge graphs, Document Classification, Information Extraction and Visualization.

Knowledge graphs: Knowledge graphs have shown to be useful in managing multi-modal data with rich metadata and relationships, most notably Google Knowledge Graph or the WikiData Graph. Also, the graph structure of a knowledge graph allows for many applications, such as advanced document retrieval [Zh18], analytics and process mining, e.g. [Ka16]. Although the benefits of having an efficient document management system have been researched [GJT19, FHP19, AA20], our approach not only aims at managing documents but to also automate document insertion and generation of rich metadata representations. There has also been research on related topics in the construction industry to represent BIM models as graphs [Fe, Is, AAW20]. Building the project management on the same technological basis will enable us to incorporate these in the future.

Document Classification: In our case, document classification is difficult for two reasons: We have to support a lot of file types and not all of them contain text. Recent works have shown success in classifying a document's type by its image representation, either multi-modal [Ru14] or also image representation alone [Fe20]. We employ the uni-modal system in [Fe20] using EfficientNet [TL19] and finetune to our task. Text classification in itself is a mature research topic. Recent advances in language modeling have shown a significant boost in classification performance, e.g. [De19, Ya19], and even proposed systems that need few task-specific data points (fewshot) or even none at all (zeroshot) [Br20, Ha20]. Zeroshot systems further have the advantage of being able to flexibly adapt to new label sets. These systems work remarkably well in single label scenarios with relatively few labels. Since in our case, text training data is very limited and labels of interest may vary, we follow [YHR19] to implement a zeroshot text classification.

[3] Video available under: `https://www.youtube.com/watch?v=DI5_go1v51k`

Information Extraction: The information we extract from the textual documents consists of entities, dates and keywords by tokenization. These are all well-researched topics and will be fully covered by Stanford's Stanza-Tool [Qi20].

Visualization: In order to access the information of the underlying document collection, visualization methods can be applied. An overview of the field is given by [Ga14]. For the visualization, we identified two main tasks for how users interact with the knowledge graph data. On the one hand, a representation of document properties and classes is needed to filter the documents. On the other hand, users also need mechanisms to detect and create connections between documents and the knowledge graph. Visualization methods focusing on the vocabulary and the general core content of documents [CCP09, St09], and also works that construct and explore relationships between documents [Wa19] are similar to our knowledge graph based approach that focuses on the named entities and the domain-specific terminology.

3 System

The system is designed to address the internal logic of a planning project in CE under German law. Additionally, we designed the system to reflect the filter and search logic familiar to engineers. Since planning bases are often delivered in large unsorted quantities (digital, analog) at the beginning of a project, a simplified data import via automatic classification and keywording was implemented. To this end, we combine the functionality of a document management data structure with NLP and knowledge representation methods. In detail, our system uses a knowledge graph and links information from unstructured data in it using document classification and information extraction. On top of the graph, we implement a visualization and navigation concept for optimized use of the data.

The backbone of the backend stack is a domain-specific knowledge graph for the construction industry in Germany in the German language. The graph contains domain-specific metadata and is used to link incoming documents with this information. Using this metadata a document is inserted into the knowledge graph and automatically linked with other documents over common edges. The knowledge graph is managed in a Neo4j[4] database and accessed by a backend server, written in Python. For easy installation and deployment, it is dockerized [Me14].

3.1 Building the Knowledge Graph

The knowledge graph covers various aspects of the German construction industry. We represent project phases, document types, participating companies as well as technical terminology and trades in this knowledge graph. To the best of our knowledge, no such

[4] www.neo4j.com

graph exists and we had to build one for the demonstrator from scratch. The necessary data was scraped from various online sources[5]. The main nodes of the graph are:

Leistungsphasen (Construction Phases): The graph contains nodes for every of the 9 "Leistungsphasen" defined in the HOAI (Honorarordnung für Architekten und Ingenieure – Fee schedule for architects and engineers), a regulation that links the fees that can be charged by engineers to construction phases and services. The phases and applicable services contained therein are the basis for the project and data organization in CE projects in Germany. The HOAI accounting scheme also contains relations like "next", "previous" and "contains" with respect to the construction phases. We map the complete structure of a project progress according to this scheme as a reference to an intended project structure.
Leistungen (construction work): Every "Leistungsphase" contains a list of basic and optional construction work, which are labeled GL and BL in the graph (German: "Grundleistung" (Basic construction work) and "Besondere Leistung" (Special construction work)). Basic services are services that are to be performed regularly within the scope of area-, object- or technical planning. They are generally required for the proper fulfillment of a construction contract and are defined by a standard. In addition to basic construction work, special construction work may be agreed upon. These construction services are not customary and not covered by any standardization and therefore not bound to the HOAI in terms of pricing.
Keywords: A keyword is an important phrase or technical term in the context of CE. In these nodes, we represent the technical vocabulary and its internal relations, which are important for the CE industry in Germany. Each node contains at least the name and a textual description as given by the scraped webpages. We use the hyperlink structure of these websites to connect these nodes. We extract the following relations from the websites: "crosslink" (description of a keyword refers to another article), "related" (the keywords are related technical terms) and "synonym" (the keywords are synonyms). For the websites containing keywords and definitions for technical vocabulary, we introduce an artificial node "Lexicon" to which all keywords are connected using the relation "part_of".
DIN/VOB: Contains DIN (Deutsches Institut für Normung – German Institute for Standardization) standards and VOB (Vergabe- und Vertragsordnung für Bauleistungen – Regulations for the awarding and contracting of construction works) regulations that are relevant to the construction industry. DIN standards are based on the verified results of science, technology and experience and serve the general public. They are developed in the process of standardization. Standards are developed with full consensus, that is, they are developed by experts to arrive at a common standpoint, taking the state of the art into consideration. The VOB is a compiled and updated three-part set of clauses for the award and contractual conditions for construction contracts. We use this standardization in CE to reference and link documents related to a standard or part of the VOB. This way, it is also possible to model relationships of documents related to the same standards or areas of construction.
Documents: These nodes manage the documents that are to be inserted into the graph when

[5] `baunormenlexikon.de/norm`, `bauprofessor.de`, `baulexikon.beuth.de`, `hausbauberater.de/fachbegriffe`

using the system. The document nodes can thus be linked to all information nodes in the graph, which allows graph search and evaluation. Relationships of the documents to project phases, technical vocabulary, trades and standards can be linked.
There are also additional node types possible, corresponding to the results of the document insertion pipeline like Persons, Locations or Organisations, which are not part of the base graph. See section subsection 3.2 for further explanation. After we create graphs for each node type based on the respective scraped web pages, they were unified into one large graph. Nodes with the same name are merged into one node. We also search in the node descriptions created by the web texts for keywords and link the node to the corresponding keyword node in case of a match. This allows us to link the standards, the HOAI and the technical terminology so that the mutual contexts and references are integrated into the knowledge graph. We further use the Wikipedia API and connect the nodes from our graph to its WikiData [VK14] ID. We search for the node title using the API and add it to the node's attributes if an unambiguous match is returned. This results in a graph with 13414 nodes and 122509 edges (0.048% density). It contains 12486 keywords, 747 norms and 64 + 47 + 58 + 9 nodes relating to the HOAI accounting scheme. 3971 nodes are connected to their corresponding WikiData ID.

3.2 Document Classification Pipeline

The core functionality of the backend is to automatically manage, organize and relate documents. The user should be able to upload any file type and the backend accumulates knowledge about its content and inserts the document into the knowledge graph. In Figure 1 the basic scheme for this process is visualized. Every document has to be automatically assigned to one of the construction phases. To be useful in a real life scenario it is important to support as many file types as possible. For files that contain text, we additionally employ the NLP pipeline.

Uploaded files are to be classified by "Leistung" (construction work). The main challenges in this regard are: At this stage of the upload the system must be able to handle (almost) arbitrary file types, not only text and PDFs, but also for example CAD, Excel, or image files and also annotated data to train supervised learning algorithms is very scarce. We address the first problem by transforming the data type to images using a thumbnailing process. Files of various types and sizes are transformed into a thumbnail using ImageMagick [Th21]. This unifies file type for classification and matches them with the networks input size. We find that this approach has the advantage to be extendable to include almost any file type whether they contain text or not. Currently, we support images and image software file types, but also many file types with textual or tabular content.[6] To address the second challenge and make efficient use of very few training data examples, we employ EfficientNet[TL19], a pre-trained model for image classification, and finetune it to classify each document with respect to the "Leistung" (construction work). Expanding this procedure to multi-modal

[6] Currently supported file types include: images, xls, xlsx, pdf, doc, docm, txt, xml, csv, html, dwt, dxf, dwg, dxf

approaches [Ru14, Fe20] combining text, image and metadata and thoroughly evaluating this process with more data will be a promising part of our future work.

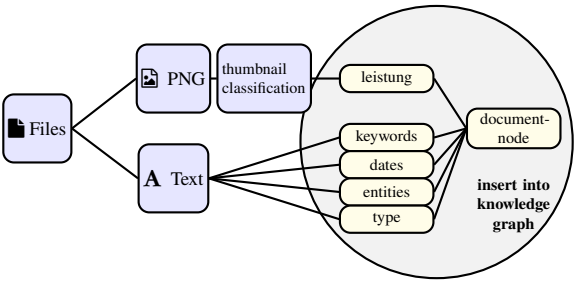


Fig. 1: Information extraction pipeline when inserting documents into the knowledge graph.

Files containing textual data are processed further: We classify the type of the document. Since we do not have annotated data regarding document type, we employ a pre-trained zeroshot model from HuggingFace [Wo20]. The German BERT [De19] model was finetuned on the German part of the XNLI [Co18] data set following the proposal of [YHR19] for zeroshot text classification. These models concatenate the textual input (premise) and the label descriptor (hypothesis) and calculate a score of how likely these entail each other. We then compare the entailment scores of all labels for the textual input and pick the one with the highest probability. This approach has three advantages: We can make reasonable predictions without training data, we can improve the classifier by collecting instances where the user corrects the prediction and update the model accordingly. Additionally, because of the premise–hypothesis mechanism, we can flexibly adapt the label set to user preference or on project basis. For our purposes, the premise consists of the textual content of the documents and the hypotheses are the candidate labels, including for example INVOICE or CONTRACT.

We extract further information from the text by searching for keywords contained in the knowledge graph (e.g. technical terms and vocabulary) and dates by regular expression. We employ the Stanza Tool [Qi20] to extract the named entities of persons, organizations and locations. All of these will also be included as nodes and edges in the graph. Documents can now be related by common semantic neighbors in the graph e.g. when both documents mention a certain time or the same person.

3.3 Graph Structure

Figure 2 shows a schematic of the graph structure after inserting documents. For simplicity only the HOAI and Lexicon sub-graph are indicated, the sub-graph is analogous for DIN/VOB nodes. The HOAI retains its original hierarchical structure in the graph and is divided into nine phases shown as $LP*$. Each of the phases consists of a list of its services, shown as $GL*$ (or $BL*$, omitted in the figure). The Lexicon is connected to all

the “keyword” nodes in the graph. Keywords can be connected with each other if a relation was either extracted from the source websites or if they contain another keyword in their own description. When inserting documents they get connected to one of the services in the graph (either GL or BL) and therefore also can be attributed to one of the phases. The extracted entities as described in the previous section are also inserted into the graph and get connected to the documents if they appear in them. Every document is also connected to every keyword that is mentioned in them. The document type classification algorithm also connects every inserted document to a document type node. Since the graph retains the hierarchical structure of the HOAI the documents can now also be accessed and sorted by project level nomenclature.

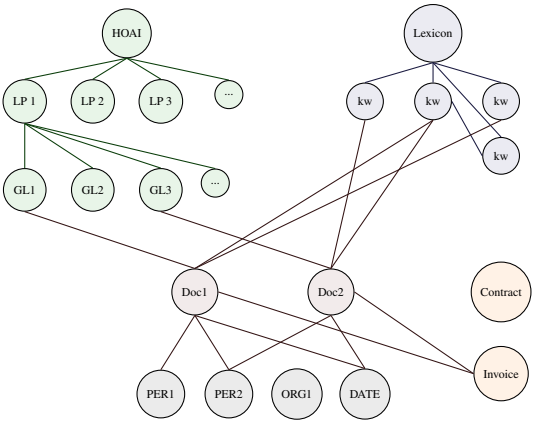


Fig. 2: Schematic of the graph structure after inserting documents.

3.4 Frontend

The frontend is implemented with React[7] running on node.js[8] as a webserver. We use d3 [BOH11] for designing the visualizations and Material UI[9] for the interface. The frontend can be seen in Figure 3 and is composed of two different views, the project overview a), and the search and context view b). It supports filtering and searching on the knowledge graph based on a tagging system, and to detect and create connections between documents and nodes.

The project overview Figure 3 a) consists of an accordion menu that shows the hierarchy of the knowledge graph and thus the project structure according to the HOAI. Starting with each “Leistungsphase” and the different “Leistungen” that are part of it. Going down the hierarchy a user can see the documents and the metadata that is part of that “Leistungsphase”

[7] https://reactjs.org/
[8] https://nodejs.org
[9] https://material-ui.com/

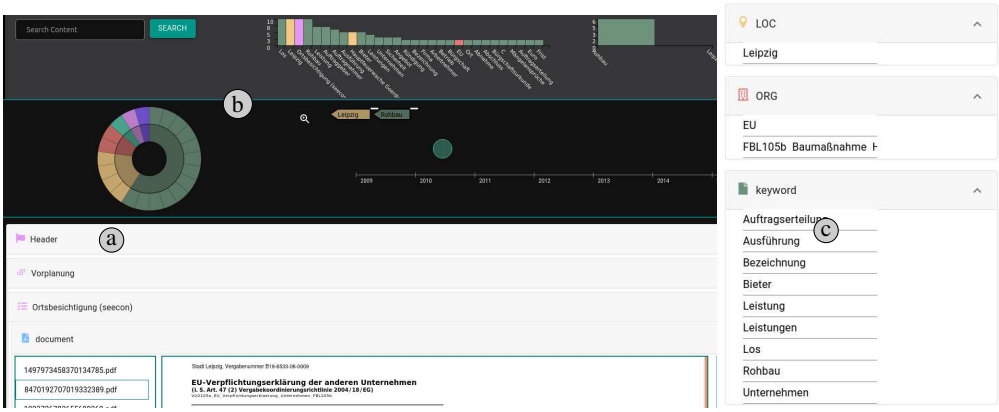


Fig. 3: The interface consisting of the project view a) and the search context view b). c) shows an increased list of the document metadata that can also be changed by the user.

or “Leistung”. For each document, the metadata that was detected based on the classification and information extraction pipeline can be seen side-by-side with the document Figure 3 c). Furthermore, for document type classification and “Leistung” classification, results that are uncertain i.e. below a specific threshold are visualized by red color inside the accordion. All of the metadata of a document can also be modified by the user e.g. to correct misclassifications and so to improve the knowledge graph. This supports the creation of new edges in the knowledge graph and also to detect connections between documents. Further, the user can upload a document collection, which is then inserted inside the knowledge graph. The accordion then shows the places the documents were inserted into. Via the menu the user can also inspect a list of all documents where classification results are uncertain and can add additional document types for the zeroshot classification.

The search and context view Figure 3 b) can be used to filter the documents through a tagging system that is based on the content and the nodes connected to the documents. A tag is added as a visual representation to the application, colored based on the node type and can be any node of the knowledge graph e.g. a specific person, a technical term or a specific construction work of a project. These tags can be added or removed to filter the data. As these tags are based on the same nodes that are displayed side-by-side with the documents a user can easily correct errors of the automated processes to ensure project consistency. Furthermore, two histograms show how the currently selected tags affect the results. The first histogram shows potential tags and how they affect the number of resulting documents to give an overview of the metadata and so reduce the potential search space. The second histogram shows how many documents are added to the search results if a specific selected tag is removed. Further, additional metadata about the search results is visualized. A timeline displays each date that is part of the selected documents as a circle. The number of occurrences of that specific date is mapped onto the size of the circles, to give an overview, which dates are frequently appearing inside the search focus. The user can

then filter based on one specific date by click on a circle or a specific time range by drag and drop interaction. Next to the timeline, a SunBurst [SZ00] shows the metadata connected to the selected documents in the search focus. The user can zoom inside the SunBurst and can interact with it in order to inspect a specific metadata category.

4 Discussion & Project Management Perspectives

With the proposed system demonstration, multiple applications are possible that can facilitate project management in the everyday life of engineers. The tool offers a perspective for automatic quality management and quality management audits. The underlying graph structure is able to not only work with multi-modal data and integrate relations between them, but also the topological features allow for the easy application of constraints and checks. For example: The data structure makes it possible to store plausibility checks as a sub-graph structure in the system and thus realize automatic quality control. Such constraints and completeness checks on the graph structure can be used to check whether all necessary documents are available and if they are assigned to the right phase of a project. For example, it can be checked whether all documents for the basic determination of a project are available so that the concrete planning can start. Additionally, red flags can be raised if an inserted document is not in the right place due to its context or is missing for a desired state of the project.

The system represents the data using a tag/label system derived directly from the knowledge graph. By adding or deleting tags via the system interface, users are able to directly and quickly re-assign the documents and apply corrections. In this way, system users have full control regarding the data structure. With some knowledge of CE's project structure, project assistants alongside engineers can thus maintain the data and keep it consistent. The assignment of tags is automated by classification components. This serves an initial assignment of a position of the data in the graph structure. Machine learning errors can also be easily corrected within the user interface. With the continuous input and correction of data, the available training data set is constantly expanded and the classification algorithms can thus be retrained. This concerns in particular the automatic organization, the sharing and finding of data. Thus, the tool simultaneously assumes a fiduciary function and can take over the data organization for larger stakeholder networks through a uniform and consistent management. Since an access matrix can be imagined here and documents no longer have to be physically transported, the security of the data is raised.

5 Conclusion

With our demonstrator, we present a new interface paradigm for managing and sharing documents in the CE sector. With it, we show perspectives for the organization of planning documentation. With an industry-compliant data structure, we optimally represent

information in the way engineers work. The results contribute a new way of thinking about collaboratively managing and sharing planning documentation with a large stakeholder network.

6 Acknowledgments

This research was funded by the Development Bank of Saxony (SAB) under project number 100335729.

Bibliography

[AA20] Ayoush, Maha; Ahmad, Hesham: Investigating the Use of Electronic Documents in the Jordanian Construction Projects. pp. 165–174. January 2020.

[AAW20] Ahankoob, Alireza; Abbasnejad, Behzad; Wong, Peter: The Support of Continuous Information Flow Through Building Information Modeling (BIM). pp. 125–137. January 2020.

[Bo21] Borst, Janos; Meinecke, Christofer; Wiegreffe, Daniel; Niekler, Andreas: , Baugraph, 2021.

[BOH11] Bostock, Michael; Ogievetsky, Vadim; Heer, Jeffrey: D^3 data-driven documents. IEEE transactions on visualization and computer graphics, 17(12):2301–2309, 2011.

[Br20] Brown, Tom; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared D.; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav; Sastry, Girish; Askell, Amanda; Agarwal, Sandhini; Herbert-Voss, Ariel; Krueger, Gretchen; Henighan, Tom; Child, Rewon; Ramesh, Aditya; Ziegler, Daniel; Wu, Jeffrey; Winter, Clemens; Hesse, Chris; Chen, Mark; Sigler, Eric; Litwin, Mateusz; Gray, Scott; Chess, Benjamin; Clark, Jack; Berner, Christopher; McCandlish, Sam; Radford, Alec; Sutskever, Ilya; Amodei, Dario: Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems, 33:1877–1901, 2020.

[CCP09] Collins, Christopher; Carpendale, Sheelagh; Penn, Gerald: Docuburst: Visualizing document content using language structure. In: Computer graphics forum. volume 28. Wiley Online Library, pp. 1039–1046, 2009.

[Co18] Conneau, Alexis; Rinott, Ruty; Lample, Guillaume; Williams, Adina; Bowman, Samuel R.; Schwenk, Holger; Stoyanov, Veselin: XNLI: Evaluating Cross-lingual Sentence Representations. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2018. event-place: Brussels, Belgium.

[De19] Devlin, J.; Chang, Ming-Wei; Lee, Kenton; Toutanova, Kristina: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In: NAACL-HLT. 2019.

[Fe] Ferguson, Holly; Ii, Charles Vardeman; Krisnadhi, Adila; Cheatham, Michelle: Semantic-Graph based approach to BIM, resilience, and spatial data interoperability. p. 27.

[Fe20] Ferrando, Javier; Domínguez, J. L.; Torres, Jordi; Garcia, Raul; Garcia, David; Garrido, Daniel; Cortada, Jordi; Valero, M.: Improving Accuracy and Speeding Up Document Image Classification Through Parallel Systems. Computational Science – ICCS 2020, 12138:387 – 400, 2020.

[FHP19] Fernando, Hansi; Hewavitharana, Thathsarani; Perera, Asoka: Evaluation of Electronic Document Management (EDM) systems for construction organizations. pp. 273–278, July 2019.

[Ga14] Gan, Qihong; Zhu, Min; Li, Mingzhao; Liang, Ting; Cao, Yu; Zhou, Baoyao: Document visualization: an overview of current research. Wiley Interdisciplinary Reviews: Computational Statistics, 6(1):19–36, 2014.

[GJT19] Guo, Fangyu; Jahren, Charles; Turkan, Yelda: Electronic Document Management Systems for the Transportation Construction Industry. International Journal of Construction Education and Research, 17:1–16, November 2019.

[Ha20] Halder, Kishaloy; Akbik, Alan; Krapac, Josip; Vollgraf, Roland: Task-Aware Representation of Sentences for Generic Text Classification. In: Proceedings of the 28th International Conference on Computational Linguistics. International Committee on Computational Linguistics, Barcelona, Spain (Online), pp. 3202–3213, 2020.

[Is] Ismail, Ali; Nahar, Ahmed; Scherer, Raimar; Dresden, TU: Application of graph databases and graph theory concepts for advanced analysing of BIM models based on IFC standard. p. 12.

[Ka16] Kalenkova, Anna A.; van der Aalst, Wil M. P.; Lomazova, Irina A.; Rubin, Vladimir A.: Process Mining Using BPMN: Relating Event Logs and Process Models. In: Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems. MODELS '16, Association for Computing Machinery, New York, NY, USA, p. 123, 2016.

[Me14] Merkel, Dirk: Docker: lightweight linux containers for consistent development and deployment. Linux journal, 2014(239):2, 2014.

[Qi20] Qi, Peng; Zhang, Yuhao; Zhang, Yuhui; Bolton, Jason; Manning, Christopher D.: Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020.

[Ru14] Rusiñol, Marçal; Frinken, Volkmar; Karatzas, Dimosthenis; Bagdanov, Andrew; Lladós, Josep: Multimodal page classification in administrative document image streams. International Journal on Document Analysis and Recognition (IJDAR), 17:331–341, December 2014.

[St09] Strobelt, Hendrik; Oelke, Daniela; Rohrdantz, Christian; Stoffel, Andreas; Keim, Daniel A; Deussen, Oliver: Document cards: A top trumps visualization for documents. IEEE transactions on visualization and computer graphics, 15(6):1145–1152, 2009.

[SZ00] Stasko, John; Zhang, Eugene: Focus+ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings. IEEE, pp. 57–65, 2000.

[Th21] The ImageMagick Development Team: , ImageMagick, January 2021.

[TL19] Tan, Mingxing; Le, Quoc: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In (Chaudhuri, Kamalika; Salakhutdinov, Ruslan, eds): Proceedings of the 36th International Conference on Machine Learning. volume 97 of Proceedings of Machine Learning Research. PMLR, pp. 6105–6114, 09–15 Jun 2019.

[VK14] Vrandečić, Denny; Krötzsch, Markus: Wikidata: A Free Collaborative Knowledgebase. Commun. ACM, 57(10):78–85, September 2014.

[Wa19] Wang, Xu; Cui, Zuowei; Jiang, Lei; Lu, Wenhuan; Li, Jie: Wordlenet: a visualization approach for relationship exploration in document collection. Tsinghua Science and Technology, 25(3):384–400, 2019.

[Wo20] Wolf, Thomas; Debut, Lysandre; Sanh, Victor; Chaumond, Julien; Delangue, Clement; Moi, Anthony; Cistac, Pierric; Rault, Tim; Louf, Rémi; Funtowicz, Morgan; Davison, Joe; Shleifer, Sam; von Platen, Patrick; Ma, Clara; Jernite, Yacine; Plu, Julien; Xu, Canwen; Scao, Teven Le; Gugger, Sylvain; Drame, Mariama; Lhoest, Quentin; Rush, Alexander M.: Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, Online, pp. 38–45, October 2020.

[Ya19] Yang, Zhilin; Dai, Zihang; Yang, Yiming; Carbonell, J.; Salakhutdinov, R.; Le, Quoc V.: XLNet: Generalized Autoregressive Pretraining for Language Understanding. In: NeurIPS. 2019.

[YHR19] Yin, Wenpeng; Hay, Jamaal; Roth, Dan: Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. In (Inui, Kentaro; Jiang, Jing; Ng, Vincent; Wan, Xiaojun, eds): Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019. Association for Computational Linguistics, pp. 3912–3921, 2019.

[Zh18] Zhang, Zhiqiang; Wang, Linan; Xie, Xiaoqin; Pan, Haiwei: A Graph Based Document Retrieval Method. pp. 426–432, May 2018.

Towards the Operationalization of Trustworthy AI: Integrating the EU Assessment List into a Procedure Model for the Development and Operation of AI-Systems

Henrik Kortum[1], Jonas Rebstadt[1], Tula Böschen[1], Pascal Meier[1], Oliver Thomas[1]

Abstract: Artificial intelligence (AI) is increasingly permeating all areas of life and not only changing coexistence in society for the better. Unfortunately, there is an increasing number of examples where AI systems show problematic behavior, such as discrimination or insufficient accuracy, missing data privacy or transparency. To counteract this trend, an EU initiative has drafted a legal framework and recommendations on how AI can be more trustworthy and comply with people's fundamental rights. However, fundamental rights are currently not reflected in procedure models for the development and operation of AI systems. Our work contributes to closing this gap so that companies, especially SMEs with small IT departments and limited financial resources, are supported in the development process. Within the framework of a structured literature review, we derive a procedure model for the development and operation of AI systems and subsequently integrate concrete recommendations for achieving trustworthiness.

Keywords: Trustworthy AI, Procedure model, Explainable AI, Machine Learning, SME

1 Motivation

Artificial intelligence (AI) and Machine Learning (ML) are permeating all areas of life, changing the way we live together in society, not only for the better [WaSi18]. There is an ever-growing number of examples in which AI systems have shown behavior that is inconsistent with fundamental rights [CaCS20, VaWi18]. For instance, they reinforce existing discriminatory biases [RoBL20], cannot meet high quality requirements in practical use [LeLL21] and omit interests of useres with respect to data privacy, transparency and autonomy. Examples include AI systems in recruiting that discriminate against females, certain ethnic groups or people with disabilities [BaHN17] or chatbots with racist, sexist, and antisemitic tendencies [WoMG17]. Moreover, the extended application of ML establishes new possibilities for cybercrime, e.g. by injecting adversarial examples using perturbations of the input vector which are in some cases not even noticeable by humans and lead to undesired behavior of the AI system [YHZL19]. In order to address those deplorable conditions, the European Union (EU) has specified regulatory requirements and developed a legal framework for the regulation of AI systems [Euro21] that is expected to have a massive impact on AI deploying and developing companies. Up until now, research in the field of ethical AI has been predominantly

[1] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH – Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, {henrik.kortum, jonas.rebstadt, tula.boeschen, pascal.meier, oliver.thomas}@dfki.de

doi:10.18420/inf2022_26

conceptual and focused on defining principles [SeBM21]. But as stated by [MiCo19] 79% of all tech workers demand specific, practical resources regarding ethical considerations. Small and medium-sized enterprises (SMEs) in particular, where the use of AI is far less widespread than in larger companies [BaDK20], will face further challenges from increasing regulation of AI. However, such regulation also offers competitive opportunities, especially for European companies [FeDe21] and SME [TiOH21]. Following this argumentation, the goal of this paper is to increase the potential for simple, guided ethical consideration by mapping the ALTAI [KoGe20] assessment questions onto a practical procedural model. In current practice, the development of AI systems is often based on the use of established procedure models, such as the Cross Industry Standard Process for Data Mining (CRISP-DM) or Knowledge Discovery in Database (KDD). However, these models do not take into account how AI systems should be designed in a trustworthy way to specifically avoid discrimination. Our work attempts to resolve this research gap by answering the following research question:

RQ: How can a procedure model for the development and operation of trustworthy AI systems be designed in accordance with the Assessment List for Trustworthy AI (ALTAI)?

Within a structured literature review, we identify relevant phases and activities from existing procedure models for the development and operation of AI systems. From the results of the literature review we subsequently derive an integrated procedure model. The resulting model is supplemented with concrete guidelines for the development of trustworthy AI systems, which we derived from the ALTAI. In doing so, we add value to both the IS research community and practitioners, providing them with the procedural model for developing and operating trustworthy AI systems. The responsible expert group of the European Commission proposes the establishment of cross-functional teams, consisting of AI developers, data scientists, procurement officers, front-end staff, legal/compliance officers and managers, for the implementation of the ALTAI. For smaller companies, however, such project stuffing hardly seems feasible, which is why we would like to explicitly address these companies with our work. The rest of our paper is structured as follows. Section 2 provides an overview of relevant literature. In Section 3 we introduce our design science-based research approach. In 4.1, we derive an integrated procedure model based on the findings of a structured literature review Section 4.2 presents the core aspects of ALTAI and provides the foundation of our requirements for trustworthy AI systems development. Subsequently, in Section 4.3, we combine the model from 4.1 with the requirements from 4.2 and propose our final model. We discuss the results in Section 5 and finalize with possible future directions in Section 6.

2 Procedure Models and Trustworthiness of AI Systems

The development of software and AI systems can be structured in typical activities and phases as is done by procedure models in classical software development. Likewise to the waterfall model [Royc87] in classical software development, there are some established

procedure models for data mining and the development of AI systems. For example, KDD (Knowledge Discovery in Databases) was one of the first models addressing the particularities of data mining [FaPS96, MaMF10]. [CCKK00] introduced CRISP-DM, which is widely used in practice It provides a standard for data mining projects and comes with relevant implications for the development of AI systems. In CRISP-DM, the essential phases "Business Understanding", "Data Understanding", "Data Preparation", "Modeling", "Evaluation" and "Deployment" are differentiated, including constant feedback between the different phases. Besides KDD as foundation for most current models and CRISP-DM already merging phases from multiple models, there is a variety of models with a holistic or a domain specific approach like [SiSS19] for the public mental health sector [MaMF10]. Alongside the demand for efficient development, the representation of ethical and regulatory implications becomes increasingly important in procedure models, especially since the EU's publication of a draft for a legal framework [Euro21].

The EU embosses the term of ethical or trustworthy AI. It summarizes a selection of concepts and principles, which constitute the robust, legally secure, and ethically defensible development and use of AI systems. Even though the EU's Guideline for Trustworthy AI initially defines its principles in an abstract manner as well, it is one of the first guidelines that offers an intuitive resource for practioneers, namely ALTAI. On an abstract level, the guideline is based on three key concepts: Lawfulness, ethics, and robustness [High18]. The ethical perspective is broken down into four principles. (1) *Respect for human autonomy* ensures that AI systems enhance, complement, and promote human capabilities without limiting their freedom or autonomy. (2) *Prevention of harm* focuses on the mental and physical integrity of humans interacting directly or indirectly with AI systems. (3) *Fairness* represents the focus of this publication and puts forward fairness and non-discrimination in the development, deployment, and use of AI systems. The aim here is to ensure equal opportunities on the one hand, but also to create the possibility of appealing against decisions made by the systems. The starting point for this is the transparency of the AI system's decision-making process, which is made explicit by the principle of (4) *explicability*. According to this principle, the purpose and capabilities of the AI system should be disclosed, and it should be possible to explain the decisions to all affected persons. The practical specification of these principles, including the resulting questions relevant for the stakeholders involved in the development process of AI-systems (ALTAI), comes with high ethical integrity for the following reasons: ALTAI was published by the Independent High-Level Expert Group on Artificial Intelligence (AI HLEG) which was set up by the European Commission. The EU submits to basic ethical values that are represented as fundamental rights. The expert group states that its motivation to encourage trustworthy AI is the compliance of fundamental rights. Members of AI HLEG come from diverse backgrounds, so that economy and science, legal authorities and different nationalities are represented in their views and interests. Also, an earlier version of the document went through a multimodal piloting phase such that broad feedback is incorporated to the final version of the assessment list. Our goal is to increase the potential for simple, guided ethical consideration by providing a mapping of ALTAI's assessment questions to a procedure model. We employ ALTAI as a basis because it

comes in a form that matches our line of thought and progress: Since ALTAI is designed as a list for self-assessment, it is framed in a user-friendly, concrete way. Additionally, its specific relevance for the European community and its high level of congruence to the principles mentioned by [JoIV19], encourage our choice. Although ALTAI is of integrity with respect to its development and design, it has to be pointed out that it does not claim to be complete. Our literature review supports that different ethical assessment lists are hardly congruent. Their focus depends on the field of application as well as the perspectives of its authors and ethical priorities that can also be contingent on time, culture and more. A selection process of headwords necessarily requires some ethical concerns to be disregarded. Also, during the application of an assessment list, trade-offs between the principles might present themselves as unavoidable. These kinds of choices reflect embeddedness of authors and developers in society, culture and other circumstance and should be continuously and critically monitored.

3 Research Approach

In order to derive an integrated procedure model for the development and operation of AI systems, a structured literature search according to [WeWa02] was carried out. Aim of the conducted literature search was the identification of relevant phases and activities of procedure models for the development and operation of AI systems. For this purpose, the following search term was applied to eight relevant databases in the field of information systems research: (*"procedure model" OR "process model" OR lifecycle OR framework) AND ("data science" OR "data mining" OR "data engineering" OR "machine learning" OR "artificial intelligence" OR "knowledge discovery" OR "decision support systems")*.The results were analyzed, based on the relevance of the publications in relation to our research question, by first analyzing the title, then the abstract and finally the full text. Tab. 1 shows the hit quantities of the individual analysis steps per database. The remaining 80 relevant publications were analyzed and categorized based on [WeWa02], identifying 7 phases and 44 relevant activities as presented in Section 4.1.

Database	Hits	On title	On abstract
IEEE	239.760	118	42
AISel	18.257	27	17
SpringerLink	176.995	82	29
Science Direct	133.311	46	21
Wiley	171.846	56	13
Ebsco	2	0	0
Web of Science	89.509	71	32
JSTOR	0	0	0
		Summed up	**154**
		Duplicates, no access or out of scope	**74**
		Total	**80**

Tab. 1: Overview: Literature review

In order to determine the trustworthiness requirements of AI services, we analyzed the ALTAI in detail (Section 4.2). The subsequent mapping of the integrated process model and ALTAI was carried out independently by three experts and consolidated within a workshop. The method of theoretical reasoning was used for this purpose. Outcome of the workshop and answer to our research question is our proposed procedure model for the development and operation of trustworthy AI systems (Section 4.3).

4 Derivation of a Procedure Model for the Development and Operation of Trustworthy AI Systems

4.1 Integration of existing Procedure Models from the Literature

As described in Section 3, we conducted an extensive search of existing models as part of a structured literature review to identify relevant phases and activities. Over the years several procedure models for AI system development have been developed. Even though the models differ in several aspects, most of them have an overlap in their core activities. It is worth noting that most of the identified models reference CRISP-DM and are guided by its core structure, which is why we also choose it as the structural starting point of our work. We divided the identified activities into seven main phases based on CRISP-DM, added additional activities and a complementary operations phase to adequately address the recommended actions derived from ALTAI later in 4.3.
The development procedure is initiated by the *Business Understanding* Phase, summing up seven activities. It begins with the activity *Explore Goal* [MCFH19], focusing on the exploration of the reason and the target of the recently initiated development project. Setting up on the defined goal, the background of the problem needs to be evaluated in the activity *Determine Business Objectives* [CCKK00], quantifying the success or the failure of the system, allowing the system to optimize towards the defined goal. As foundation for the project plan, in the *Assess Situation* [CCKK00] activity, resources, costs and benefits are contemplated. To consider possible risks, the activity *Identify Risk Factors* [SiSS19] has been defined, followed by the three final requirement regarded activities: Initially, one needs to *Elicitate Requirements* [ANIN17], followed by *the Requirement Negotiation* [ANIN17] to fix the requirements in a unanimous state and finalized by the *Requirement Validation* [ANIN17]. Building up on the validated requirements, the second phase is *Data Understanding*, the first of two explicitly data-related phases. As initial activity, called *Acquire Data* [MCFH19], all available data is collected. This data needs to be explored in the *Explore Data* activity [MCFH19], described in the *Describe Data* [CCKK00] activity and quality-checked in the activity *Verify Data Quality* [CCKK00]. After the technical evaluation of the data the goal-specific value of the data is quantified in the *Data Value Exploration* [MCFH19]. As foundation for the *Data Preparation* valuable stories are extracted from the data in the *Narrative Exploration* [MCFH19]. In the second explicitly data related phase, *Data Preparation*, the data is set up for the training procedure. First, the relevant data is selected (*Select Relevant Data* [LaRa11]),

cleaned (*Clean Data* [CCKK00]) and variables can be engineered (*Construct Data* [CCKK00]). In case of lacking or unbalanced training data, new data should be generated [MCFH19]. According to task-specific requirements, the resulting data is formalized in the *Formalization of Data* activity and a data architecture is defined (*Define Data Architecture* [MCFH19]), to physically and logically structure the data sources. As final steps, the data is integrated in a single structure (*Integrate Data* [CCKK00]) and formatted into the required structure in the *Format Data* [CCKK00] activity. In the following *Modeling* Phase, the model is developed, starting with the first activity, called *Select Modeling Techniques* [CCKK00] and guarded by the development of a corresponding test design (*Generate Test Design* [CCKK00]). Building up on this, the actual model is developed in the activity *Build Model* [CCKK00]. The model should be compared to an own, task-specific baseline model (*Comparing to Baseline Model* [Wang08]) and assessed in general (*Assess Model* [CCKK00]). As final activity of the current phase, explanatory methods are added, if the model itself is not already inherently interpretable as can be the case in (learned) decision trees or rule-based systems [SaGr17]. To assess and verify the quality of the developed model, the *Evaluation* phase is carried out. The evaluation is done in three main activities, starting with an exploration (*Explore Results* [MCFH19]) and an evaluation (*Evaluate Results* [CCKK00]) of the model-results to address all technical concerns. Next to this, the conducted process needs to be reviewed as well in an activity called *Review Process* [CCKK00]. As one of the most extensive phases in the development of productive AI systems, the *Deployment* contains all activities from the evaluated model towards a releasable system. As initial activity, setting up on the evaluation results, in the *Explore Product Opportunities* activity [MCFH19], the added value and existing possibilities of the developed model in a product setting are assessed. If the development of a product is promising, the activity *Plan Deployment* [CCKK00] is started, followed by the release of the data in the future production environment (*Release Data* [MCFH19]). Using this and the predefined model, the system is implemented (*Implement System* [HSMK19]), incorporating the development of a User Interface (*Implement User Interface (UI)* [LaRa11]) and the inclusion of the developed explanatory approaches (*Integrate Explanations in UI* [SaGr17]). The resulting system needs to be tested as a whole, extending the model-specific evaluation (*Test System* [HSMK19]). After the test procedure, a final report is written (*Produce final Report* [CCKK00]), the system is released (*Release System* [HSMK19]) and finally, the project needs to be reviewed (*Review Project* [CCKK00]). As extension to the main phases of CRISP-DM, the Operation phase has been added in this model, to continuously supervise and optimize the running AI system and incorporate all user related communicatory topics. As initial only activity, *Monitor System* [HSMK19] is focusing on the continuous behavior of the developed system not only once after the development but rather during the whole life cycle of the system. In the following section our proposed model is supplemented by concrete recommendations for action resulting from ALTAI.

4.2 Derivation of a Procedure Model for the Development and Operation of Trustworthy AI Systems

Based on the considerations outlined in Section 2, we decided to use ALTAI as the basis for the assessment regarding trustworthiness, which we include in the procedure model. ALTAI addresses AI developers, project managers, front-end staff and more professionals that work with AI systems. The assessment list has seven sections each of which present a number of headwords. Every headword comes with a set of questions for self-assessment. The seven sections and headwords are presented in Tab. 2. The headwords are named in the brackets and referenced by an identifier. In Section 4.3 we will assign the individual headwords of ALTAI to the appropriate phases and activities of the procedure model in which they are relevant and briefly explain how they can be addressed.

ID	Section Name	Summary and Headwords
R1	Human Agency and Oversight	▪ Assess possible influences of the AI system to individual humans, particularly as the systems guides, influences or supports human decision making (Human Agency and Autonomy, #1.1) ▪ Enable humans to intervene AI system at all times (Human Oversight, #1.2)
R2	Technical Robustness and Safety	▪ Protect the system from physical and cyber-attacks and assess the risks that arise in case of abuse/deficiency (Resilience to Attack and Security, #2.1) ▪ Assess risks that might arise from sloppy design (General Safety, #2.2) ▪ Assess the effects that inaccurate predictions of the system would put forward (Accuracy, #2.3) ▪ Put forward means to compensate for the system in case of failure and ongoingly validate it (Reliability, fallback plans and reproducibility, #2.4)
R3	Privacy and Data Governance	▪ Handle personal (user) data responsibly (Privacy, #3.1) ▪ Assure integrity of data quality and content (Data Governance, #3.2)
R4	Transparency	▪ Assure that the principle of operation and the decisions of the AI system remain traceable (Traceability, #4.1) ▪ Encourage the user's understanding of the AI system's decisions (Explainability, #4.2) ▪ Communicate possible risks and limitations of the AI system to users and, if applicable, provide disclaimers (Communication, #4.3)
R5	Diversity, Non-Discrimination and Fairness	▪ Design data sets and algorithms such that results are fair with respect to diversity and representativeness (Avoidance of unfair bias, #5.1) ▪ Make sure that the system can be used by everyone, including people with special needs or preferences (Accessibility and Universal Design, #5.2) ▪ Consult stakeholders during the development of the AI system (Stakeholder Participation, #5.3)
R6	Societal and Environmental Well-Being	▪ Monitor and reduce negative impacts on the environment (Environmental Well-Being, #6.1) ▪ Monitor impact on the working environment and required skills and make efforts to adapt (Impact on Work and Skills, #6.2) ▪ Monitor and reduce negative impact that the AI system may have on society and democracy (Impact on Society at large or Democracy, (#6.3)
R7	Accountability	▪ Make sure that the system can be audited independently from its development (Auditability, #7.1) ▪ Constantly monitor possible risks that arise in the scope of the AI system and explicate trade-offs between the ethical principles (Risk Management, #7.2)

Tab. 2: Summary of ALTAI [KoGe20]

4.3 Extension of the Procedure Model to include Trustworthiness

In the next step, the aggregated model (see Section 4.1) is enriched by the ALTAI recommendations for action (see Section 4.2), which are assigned to the appropriate phases and activities. It is worth noting that not all requirements derived from ALTAI could be assigned reasonably to any of the activities identified in the literature, which is why we added three new activities to the model. Fig. 1 shows our final procedure model for the development and operation of trustworthy AI systems, with the individual recommendations for actions resulting from the ALTAI grouped by phase described in detail below.

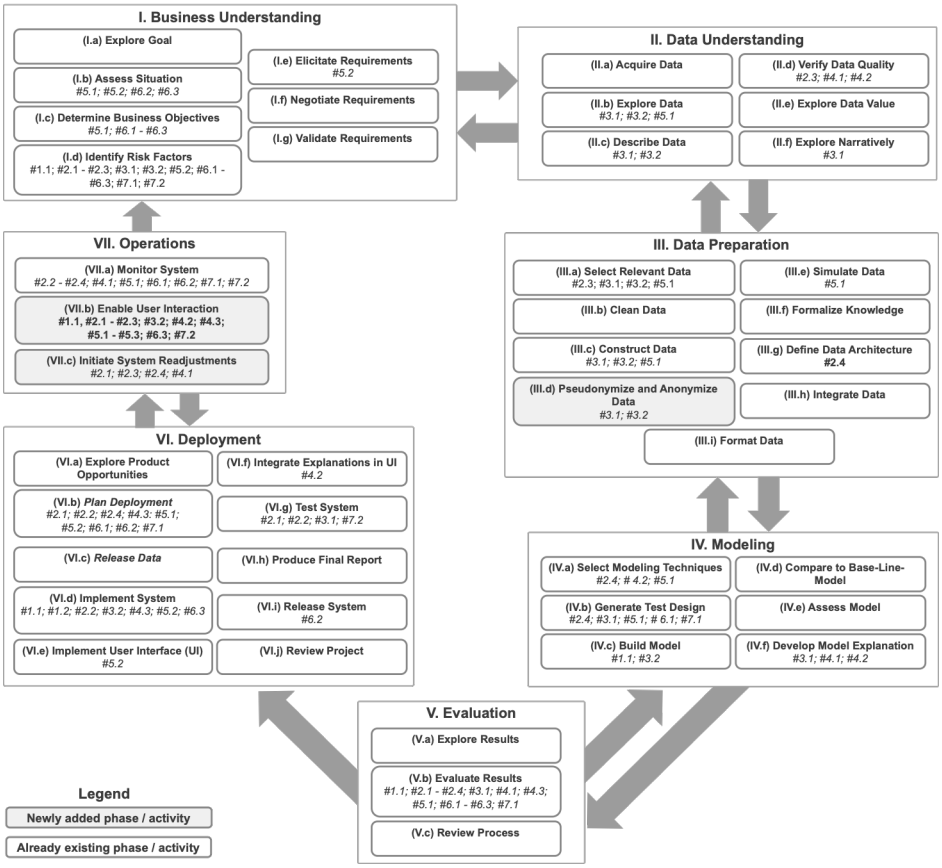


Fig. 1: Procedure model for the development and operation of trustworthy AI-systems

(I) Business understanding: In this phase, requirements of all categories must be considered. When assessing the situation, specific attention should be paid to identify potential biases and to understand the diversity and representativeness of the end-users (#5.1). Even beyond the end users, other affected stakeholders should also be identified

and involved in this activity to be aware of potential discrimination and to include the needs of all stakeholders (#5.2). Of particular importance here are also the workers interacting with the system. AI changes the demands on the labor market [KoRT22], therefore, the potential impact on the skillsets required for the tasks and on work arrangements must be determined (#6.2). It should also be evaluated whether the AI system could have a potential impact on society (#6.3). The next relevant activity, the identification of risk factors, is of central importance for the development of trustworthy AI systems. The relevant risk factors are distributed across all categories of the ALTAI. They include the assessment of possible influences of the AI system on individual humans, especially when the system is guiding human decision making (#1.1). Furthermore, an assessment of the impact caused by misuse and attacks on the AI system (#1.1), sloppy design (#2.2) and inaccurate predictions (#2.3) of the AI system must be made here. Another potential source of risk is inadequate privacy (#3.1) and data protection (#3.2) measures. The risk and consequences of possible inequitable behavior of the system toward end users or individual subgroups must also be assessed (#5.2). In addition to the concrete risks with direct effects, long-term risks with potentially negative consequences for the environment (e.g., due to non-environmentally sustainable operation) (#6.1) and for the working environment (#6.2) of users, and more abstractly, for society as a whole (#6.3) need to be addressed. All these aspects should be ascertainable in a structured way by means of an audit mechanism (#7.1). For organizational guidance, ALTAI recommends third-party consulting, establishing an AI ethics review committee, and providing legal and risk trainings (#7.2). Based on the risk factors, concrete requirements should be identified, negotiated and evaluated. Here, the aspect of accessibility and universal design (#5.2) is decisive. The requirements should correspond to the diversity of preferences and capabilities in society. Special care should be taken to ensure that accessibility to the AI system for persons with disabilities, is adequately addressed in the requirements.

(II) Data understanding: During the exploration of the data, care should be taken to ensure that the data is appropriate and representative of the diversity of the population. In particular, attention should be paid to any bias that may exist that could result in discrimination against certain subgroups (#5.1). Building understanding about the data should be supported by publicly available technical tools. In addition, privacy (#3.1) and data governance (#3.2) aspects must also be considered during this activity. It is necessary to identify data that fall under the Data Protection Regulation (GDPR), or a non-European equivalent, or whose further processing could be critical for other reasons. Therefore, ALTAI recommends that a Data Protection Impact Assessment (DPIA) is carried out here. To allow the classification of critical data to be considered in the further process, the data should be described accordingly. The data quality assessment must ensure that data used for the AI model development is up to date, of high quality, complete and representative for the environment in which it is to be used (#2.3). To this end, ALTAI calls for the establishment of continuous measures (#4.1).

(III) Data preparation: During this phase, privacy (#3.1) and data governance (#3.2) aspects have to be considered. This includes, for example, access to data, which should be limited to authorized and qualified personnel (#3.1), and the integration and establishment of privacy-by-design measures. Only sources that comply with data protection laws or for

which the data owner has given consent should be selected for further processing (#3.1, #3.2). To comply with data governance, continuous monitoring measures should be defined and applied when constructing new data. Care should also be taken to ensure that the data selected is not biased and therefore has the potential to discriminate (#5.1). To achieve privacy-by-design and default, ALTAI recommends encryption, pseudonymization, aggregation and anonymization. Therefore, we propose the extension of the preprocessing phase to include a concrete processing activity for pseudonymization and anonymization. Care should also be taken to ensure that the data selected is not biased (#5.1). The same applies to subsequent activities where data are further processed. To avoid bias, continuous monitoring and measures such as simulation of new training data points for underrepresented classes should be included.

(IV) Modeling: Avoiding bias and the unfair discrimination of minorities (#5.1) is also a crucial requirement to be considered in the modeling phase. In concrete, for example, it should be checked whether a potentially suitable algorithm can deal well with unequally distributed data. When developing a test design, a testing activity that checks for discrimination should be explicitly integrated. For classification problems, for example, the use of the F1 difference - which relates the F1 score of the individual subclasses - is suitable [KFRE22]. Reproducibility and reliability also play an important role in modelling (#2.4). These factors must already be taken into account during model selection and integrated into a test concept. In addition to identify potentially harmful results generated by the AI model, it should also ensure that the training and execution of AI models is as resource-efficient and environmentally friendly as possible (#6.1). One measure for implementation could be, to consider the computational efficiency in the test design in addition to the pure quality of the results. Once the model training is complete, ensuring traceability (#4.1) and explainability (#4.2) of the model is important. Tracking should be provided of what data was used by the AI system to make a particular decision or in case of multiple models (stacked/ensemble architectures), which AI model or rules led to the results (#4.1). This property is closely related to the characteristic of model explainability. It should be ensured that both the technical processes of the AI system are explainable transparent as well as the reasoning behind the predictions of the AI algorithm.

(V) Evaluation: During evaluation the results of the model must be examined. Here, a variety of aspects and dimensions of ALTAI must be considered, as many of the aspects previously defined as requirements can be concretely evaluated based on the model results. For example, it should be specifically checked for discrimination using the previously defined performance metrics (#5.1). Another important aspect of the evaluation is to check the resilience of the AI model, since AI algorithms offer entirely new attack vectors. E.g. deep neural networks (DNNs) are vulnerable to adversarial examples added to the input by adding perturbations not noticeable by humans (#2.1). The evaluation should cover tests that check the behavior of the algorithm precisely for such cases.

(VI) Deployment: The application embedding the AI algorithm must be reliable, deliver reproducible results and include fallback scenarios (#2.4) and allow auditing by internal and external auditors (#7.1). Specifically, failsafe scenarios should be defined that are triggered, for example, in the case of results with low confidence or obvious errors. The system should be transparent to the end user that it uses an AI algorithm. This includes

information on the purpose, added values and risks and limitations of the AI service for the specific use case (#4.3). During development, potential external attacks on the AI system must also be considered and appropriate resilience mechanisms implemented (#2.1). Here, conformity to existing standards for cyber security and, if applicable, certification should be taken into account. This goes hand in hand with the consideration of general safety features. For example, the service should be designed to be resilient by using duplication (#2.2). Creating accessibility and adhering to universal design principles are central pillars of the deployment phase (#5.2). The general design of the AI system must correspond to the diversity of preferences and abilities in society. It is important to consider user groups with special needs and disabilities. Specifically, for example, it should be ensured that the interface is also accessible by users of assistive technologies. Users should be involved in a consultative manner throughout the implementation phase to ensure that the development team has an accurate understanding of their needs expectations and potential disadvantages caused by the system, e.g. explainability is often a crucial factor in increasing user confidence in AI systems (#4.2). When developing the user interface, an explanation function should be integrated that provides the user with as much information as possible about the decisions made by the system. If possible, this can be an explanation of why a model generated a particular output or decision and what combination of input factors contributed to it. The AI system should have minimal energy consumption and carbon emissions (#6.1) to ensure the well-being of the environment. A proper implementation should be based on the existing skills of the using employees, correspond to their knowledge and include self-explanatory functions (#6.2).

(VII) Operation: During operation, ongoing monitoring should be carried out to verify that the assumptions and database on which the modeling and training were founded are valid in practice and that the requirements for AI accuracy are met (#2.3). Such monitoring also serves to check the reliability (#2.4) and traceability (#4.1) of the AI system. A poor performance of the model can lead to unintended biases and thus to discrimination against individual groups (#5.1). To detect and prevent such misbehavior at an early stage, the previously selected fairness metrics should be continuously surveyed. In addition, the monitoring should also cover the resource consumption incurred for the model's operation (#6.1). If the AI system is to be used in a work context, the impact on employees and work processes must be recorded (#6.2). All monitoring aspects should be implemented in a way that promotes the auditability of the system (#7.1). Since ALTAI goes beyond mere monitoring and explicitly call for a feedback loop and adjustment of the AI system, we suggest extending the operation phase by the activity of a readjustment. In principle, all model behavior identified while monitoring and classified as undesirable should lead to an adjustment. Here the use of online learning, i.e., adjusting the model based on live data collected in the field, allows for better fitting the real data, but should take into account possible negative consequences in terms of the AI system learning unusual or undesirable patterns (#2.4). In addition to adjusting the AI algorithm, technical service updates may also be necessary to ensure the secure trustworthy execution, e.g., in terms of resistance to malicious attacks (#2.1). User interaction and communication are key to building trust in the AI system and a central component of ALTAI. Hence, we suggest the introduction of an activity that enables user interaction before, during and after the launch of the AI

system. This starts with simple communication about the duration of security coverage and updates (#2.1), but also includes informing users about the advantages potential risks associated with its use (#2.2, #4.3) and about the quality and accuracy of the predictions provided by it (#2.3). Enabling user interaction can help to increase their autonomy by making the origin of the systems output and functioning transparent thus promoting a more conscious decision-making (#1.1) This activity involves more than one-way communication, but also provides for the recording of and the response to user feedback. For example, the user must always have the option of withdrawing consent to the use of his or her data (#3.2). In this activity, measures should also be defined to ensure that the user understands. The success of these measures should then be evaluated (#4.2). In the case that the AI system itself communicates with the user (e.g. chatbots), it must be made transparent that this is an AI interacting with the user (#4.3). In addition, training material should be provided to enable users to adequately handle the system (#4.3). The material must also be usable by people with disabilities and by users of assistive technologies (#5.2). The user must also be given the opportunity to report observations related to bias, discrimination, or poor performance of the AI system (#5.1). In general, ALTAI advises broad inclusion of all potentially relevant stakeholder groups (#5.3), also during operation.

5 Discussion

While the EU guidelines itself are rather vague and make it difficult for companies to transfer them into practice, ALTAI offers concrete guiding questions and thus a stronger practical relevance. However, there is a lack of integration into practice-oriented procedure models. The model we have proposed is intended to remedy this situation and to help small and medium-sized enterprises in particular to implement ethical AI in practice. The mapping between activities and phases of the procedure model and ALTAI requirements ensures that the relevant guidelines are considered at all times in the development process. It can be noted that the ALTAI focuses on human users and their needs and rights. The perspective is also reflected in our process model and has an influence on how companies should develop AI models in the future to enable the realization of trustworthy solutions. Another relevant aspect is the continuity of measures required by the ALTAI in many places. Its implementation through mechanisms, such as MLOps, may lead to higher one-time investment costs for SMEs, but over time it can lead to the realization of efficiency gains and economies of scale that reduce the costs per AI project.

During the development of the procedure model, we had to decide on the level of detail of the model. The model had to be general enough to be independent of the industry and use case, and specific enough to ensure that the key activities were carried out. To ensure applicability, especially for SMEs with limited resources, the model should also be pragmatic and compact. Consequently we based it on the level of detail of CRISP-DM, since it is already widely used in practice and should also be familiar to SMEs. However, since our process model is an aggregate of various models, the implications can also be applied to them. This has the potential to simplify the integration of ALTAI, especially for

practitioners who already use procedure models and cannot easily switch from one to the other. Independent mapping by three experts and subsequent alignment is intended to strengthen reliability. To further increase the validity of the model, we want to encourage experts to apply and evaluate the model in practice. By developing the process model and integrating ALTAI, we answer the RQ with an instantiation of a procedure model that enables further research projects to integrate ALTAI requirements in the development process.

Every AI project is different and therefore, in practice, the focus on certain aspects of ALTAI must also be re-evaluated depending on the use case. Currently, our model focuses on stand-alone AI systems, and on systems in which AI is a core component. In the discipline of Information Systems, our work can be assigned to the development of prescriptive knowledge and can be used in Design Science research that intends to develop AI systems according to the EU guidelines. It also allows to draw conclusions about the completeness and suitability of existing AI systems for an integration of ethical requirements. We also found that existing and established models such as CRISP-DM do not sufficiently take ethical aspects into account by themselves and even existing extensions like Rebstadt et al. [RKGE22] are only addressing selected subparts like non-discrimination. Consequently, we suggest an expansion of the present models. In practice, our model created provides users with an introduction to compliance with the ethical guidelines, which in the future will become legally binding as a result of the EU's ambitions in the Artificial Intelligence Act and could result in penalties for companies that violate them. In this way, we want to reduce hurdles and uncertainties in the adoption of AI, especially for SMEs with limited resources. Beyond this introduction, the model can be further developed both scientifically and in practice. In addition to detailing the phases, it is also possible to expand the model to include additional best practices and tools in a general or domain specific manner. For example, we see great potential in embedding our process model in a software tool to enable SMEs in particular to use it even more concretely. Such a tool could be designed as an intelligent recommendation system that queries the user for information about the project over the individual phases and, on the basis of the input, makes concrete recommendations and points out necessary actions.

To date, AI is increasingly finding its way into very private areas and determines many decisions in people' s daily lives. At the same time taking ethical factors into account and creating trustworthy AI systems already goes far beyond mere compliance with laws such as GDPR and are likely to become even more important in the future. Although the ALTAI suggests performance tests - e.g. in terms of non-discrimination - they are rather vague regarding their concrete design. In our view, this results in a need for auditing and certification of AI systems based on standardized tests. Our work also has implications for the research field around data ecosystems. On the one hand, because AI systems are playing an increasingly important role here, and on the other hand, because data sharing, which is so central to data ecosystems, involves people. The data owners and producers are often private individuals whose willingness to share data with third parties is crucial to the success of ecosystems. By taking their interests into account more strongly during the development process and ensuring that their data is processed in trustworthy services, precisely this willingness could be strengthened.

6 Conclusion and Future Work

The increasing importance of developing and operating AI systems requires that fundamental rights are considered in the development process. To help developers of AI applications comply with the high ethical requirements, we have developed a procedure model that considers the EU guidelines for trustworthy AI applications. In a first step, we analyzed the existing procedure models in the area of ML around CRISP-DM and transferred them into an integrated model. Finally, by mapping ALTAI to the integrated model, we answer our research question and present our procedure model for a trustworthy development and operation of AI systems. Our model offers users an introduction to the holistic integration of measures to avoid discrimination, insufficient accuracy, lack of data protection and transparency in the development of AI systems and thus makes an important contribution to increasing trust in AI. Our proposed procedure model represents a basic building block, which will have to be extended in the future by more detailed investigations of the individual phases and methods. It offers SMEs in particular a starting point for the ethical use of AI. The intensive consideration of ethics in the development process and during the operation of AI systems can ensure that the great potential of AI applications can be used with reduced negative side effects in organization, private life and society. Future research in this area should continue the DSR cycle and focus on further evaluation of the model in practice and by experts (such as the High-Level Expert Group on AI). With this in mind, an evaluation by SMEs from various domains is planned for the future. In this context, we would also like to encourage other researchers and practitioners to apply the model to concrete AI projects and to contribute to its further development.

Bibliography

[ANIN17] ALTARTURI, HAMZA HUSSEIN ; NG, KENG-YAP ; IZUAN, MOHD ; NINGGAL, HAFEZ ; SHAHREL, AZREE ; NAZRI, AHMAD ; AZIM, ABDUL ; GHANI, ABD: A Requirement Engineering Model for Big Data Software. In: *Conference on Big Data and Analytics (ICBDA*, 2017 — ISBN 9781538607909, S. 111–117.

[BaDK20] BAUER, MARKUS ; VAN DINTHER, CLEMENS ; KIEFER, DANIEL: Machine learning in SME: an empirical study on enablers and success factors (2020).

[BaHN17] BAROCAS, SOLON ; HARDT, MORITZ ; NARAYANAN, ARVIND: Fairness in machine learning. In: *Nips tutorial* Bd. 1 (2017), S. 2017.

[CaCS20] CASTILLO, DANIELA ; CANHOTO, ANA ISABEL ; SAID, EMANUEL: The dark side of AI-powered service interactions: Exploring the process of co-destruction from the customer perspective. In: *The Service Industries Journal*, Taylor & Francis (2020), S. 1–26.

[CCKK00] CHAPMAN, PETE ; CLINTON, JULIAN ; KERBER, RANDY ; KHABAZA, THOMAS; REINARTZ, THOMAS ; SHEARER, COLIN ; WIRTH, RÜDIGER: CRISP-DM 1.0: Step-by-step data mining guide Bd. 16 (2000), S. 1–76.

[Euro21] EUROPEAN UNION, EU: Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS, 2021.

[FaPS96] FAYYAD, USAMA ; PIATETSKY-SHAPIRO, GREGORY ; SMYTH, PADHRAIC: The KDD Process for Extracting Useful Knowledge from Volumes of Data. In: *Communications of the ACM* Bd. 39 (1996), Nr. 11, S. 27–34.

[FeDe21] FERRÁNDIZ, ESTER MOCHOLÍ ; DEGLI-ESPOSTI, SARA: After the GDPR: Cybersecurity is the Elephant in the Artificial Intelligence Room. In: *European Business Law Review* Bd. 32 (2021), Nr. 1.

[High18] HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, EUROPEAN COMMISSION: *ETHICS GUIDELINES FOR TRUSTWORTHY AI*. Brüssel, 2018.

[HSMK19] HESENIUS, MARC ; SCHWENZFEIER, NILS ; MEYER, OLE ; KOOP, WILHELM ; GRUHN, VOLKER: Towards a software engineering process for developing data-driven applications. In: *Proceedings - 2019 IEEE/ACM 7th International Workshop on Realizing Artificial Intelligence Synergies in Software Engineering, RAISE 2019*, IEEE (2019), S. 35–41 — ISBN 9781728122724.

[JoIV19] JOBIN, ANNA ; IENCA, MARCELLO ; VAYENA, EFFY: Artificial Intelligence: the global landscape of ethics guidelines (2019).

[KFRE22] KORTUM, HENRIK ; FUKAS, PHILIPP ; REBSTADT, JONAS ; ELEKS, MARIAN ; NOBAKHT GALEHPARDSARI, MARJAN ; THOMAS, OLIVER: Proposing a Roadmap for Designing Non-Discriminatory ML Services: Preliminary Results from a Design Science Research Project (2022).

[KoGe20] KOMMISSION, EUROPÄISCHE ; GENERALDIREKTION KOMMUNIKATIONSNETZE, INHALTE UND TECHNOLOGIEN: *The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment* : Publications Office, 2020.

[KoRT22] KORTUM, HENRIK ; REBSTADT, JONAS ; THOMAS, OLIVER: Dissection of AI Job Advertisements: A Text Mining-based Analysis of Employee Skills in the Disciplines Computer Vision and Natural Language Processing. In: *Proceedings of the 55th Hawaii International Conference on System*

Sciences, 2022.

[LaRa11] LAKSHMI, B. N. ; RAGHUNANDHAN, G. H.: A conceptual overview of data mining. In: *Proceedings of National Conference on Innovations in Emerging Technology, NCOIET'11*, IEEE (2011), S. 27–32 — ISBN 9781612848082.

[LeLL21] LEBOVITZ, SARAH ; LEVINA, NATALIA ; LIFSHITZ-ASSAF, HILA: Is AI ground truth really "true"? The dangers of training and evaluating AI tools based on experts' know-what. In: *Management Information Systems Quarterly* (2021).

[MaMF10] MARISCAL, GONZALO ; MARBAN, OSCAR ; FERNANDEZ, COVADONGA: A survey of data mining and knowledge discovery process models and methodologies. In: *The Knowledge Engineering Review* Bd. 25, Cambridge University Press (2010), Nr. 2, S. 137–166.

[MCFH19] MARTINEZ-PLUMED, FERNANDO ; CONTRERAS-OCHANDO, LIDIA ; FERRI, CESAR ; HERNANDEZ ORALLO, JOSE ; KULL, MEELIS ; LACHICHE, NICOLAS ; RAMIREZ QUINTANA, MARIA JOSE ; FLACH, PETER A.: CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. In: *IEEE Transactions on Knowledge and Data Engineering* Bd. 4347 (2019), Nr. c, S. 1–1.

[MiCo19] MILLER, C ; COLDICOTT, R: People, power and technology: The tech workers' view. In: *Retrieved from Doteveryone website: https://doteveryone. org. uk/report/workersview* (2019).

[RKGE22] REBSTADT, JONAS ; KORTUM, HENRIK ; GRAVEMEIER, LAURA SOPHIE ; EBERHARDT, BIRGID ; THOMAS, OLIVER: Non-Discrimination-by-Design: Handlungsempfehlungen für die Entwicklung von vertrauenswürdigen KI-Services. In: *HMD Praxis der Wirtschaftsinformatik* Bd. 59 (2022), Nr. 2, S. 495–511.

[RoBL20] ROBERT JR, LIONEL P ; BANSAL, GAURAV ; LUTGE, CHRISTOPH: ICIS 2019 SIGHCI workshop panel report: human--computer interaction challenges and opportunities for fair, trustworthy and ethical artificial intelligence. In: *AIS Transactions on Human-Computer Interaction* Bd. 12 (2020), Nr. 2, S. 96–108.

[Royc87] ROYCE, WINSTON W: Managing the development of large software systems: concepts and techniques. In: *Proceedings of the 9th international conference on Software Engineering. IEEE Computer Society Press* (1987).

[SaGr17] SALTZ, JEFFREY S. ; GRADY, NANCY W.: The ambiguity of data science team roles and the need for a data science workforce framework. In: *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017* Bd. 2018-Janua (2017), S. 2355–2361 — ISBN 9781538627143.

[SeBM21] SEPPÄLÄ, AKSELI ; BIRKSTEDT, TEEMU ; MÄNTYMÄKI, MATTI: From Ethical AI Principles to Governed AI. In: *ICIS 2021 Proceedings* (2021).

[SiSS19] SILVA, CHARITH ; SARAEE, MAHSA ; SARAEE, MO: Data Science in Public Mental Health: A New Analytic Framework. In: *Proceedings - IEEE Symposium on Computers and Communications* Bd. 2019-June (2019), S. 1123–1128 — ISBN 9781728129990.

[TiOH21] TIMAN, TJERK ; VAN OIRSOUW, CHARLOTTE ; HOEKSTRA, MARISSA: The Role of Data Regulation in Shaping AI: An Overview of Challenges and Recommendations for SMEs. In: *The Elements of Big Data Value*, Springer, Cham (2021), S. 355.

[VaWi18] VANDERELST, DIETER ; WINFIELD, ALAN: The dark side of ethical robots. In: *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018, S. 317–322.

[Wang08] WANG, XIAOH: The realization of knowledge discovery in total quality management system. In: *Proceedings - 5th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2008* Bd. 4 (2008), S. 643–646 — ISBN 9780769533056.

[WaSi18] WANG, WEIYU ; SIAU, KENG: Ethical and moral issues with AI: a case study on healthcare robots. In: *Twenty-fourth Americas conference on information systems. Retrieved from July*. Bd. 16, 2018, S. 2019.

[WeWa02] WEBSTER, JANE ; WATSON, RICHARD T: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: *Management Information Systems Quarterly* Bd. 26 (2002), Nr. 2, S. xiii–xxiii — ISBN 02767783.

[WoMG17] WOLF, M J ; MILLER, K W ; GRODZINSKY, F S: Why We Should Have Seen That Coming: Comments on Microsoft's Tay "Experiment," and Wider Implications. In: *The ORBIT Journal* Bd. 1 (2017), Nr. 2, S. 1–12.

[YHZL19] YUAN, XIAOYONG ; HE, PAN ; ZHU, QILE ; LI, XIAOLIN: Adversarial examples: Attacks and defenses for deep learning. In: *IEEE transactions on neural networks and learning systems* Bd. 30, IEEE (2019), Nr. 9, S. 2805–2824.

AIDA-Vis
Automatic Data Visualization with Human Preferences

Walter Laurito[1], Jacqueline Höllig[1], Jonas Lachowitzer[2], Steffen Thoma[1], Matthias Budde[2], Patrick Philipp[3]

Abstract: Data visualization is a complex task that typically requires human expertise, acquired through a large number of professional working hours. The automatic generation of reasonable visualizations would be a good solution for inexperienced laypeople. However, existing approaches fall short since they are quite static and rely only on traditional supervised learning. This results in models which recommend a single visualization solely based on the dataset features. User preferences and goals are not taken into account. We propose a more flexible solution that is iteratively updated with the individual user's preferences and outputs a ranked list of visualizations for a given dataset.

Keywords: Visualization Recommendation, Automated Visualization Design, Machine Learning, Human Preferences, Reinforcement Learning, Evolutionary Algorithm, Reward Learning

1 Introduction and Related Work

Data visualizations are central for understanding, and presenting content as well as relationships in datasets. However, designing good visualizations requires a considerable amount of prior experience, data preparation, and a sense of aesthetics. Things get more complicated as new types of visualizations emerge, user demands grow, and datasets become increasingly complex. Automating parts of the data visualization process can address this complexity.

Rule-based systems rely on heuristics to generate visualizations automatically [MHS07; Wo16]. However, capturing and formalizing the complete knowledge of a good visualization design as a fixed rule set would require a considerable amount of effort. Furthermore, frequent changes could be necessary, e.g., for new forms of visualizations, differences in aesthetic preferences or to adapt to other cultural contexts. Learning-based visualization recommenders as in [DD19; Hu19; Lu18; Mo19] employ supervised learning: Dataset features are modeled, and a neural network predicts the corresponding visualization descriptions. This has the advantage that the quality of predictions increases with the number of training examples. However, these approaches only recommend a single, data-set dependent visualization description, which is problematic in practice since visualizations

[1] FZI Forschungszentrum Informatik, Haid-und-Neu-Straße 10-14, Karlsruhe, {laurito,hoellig,thoma}@fzi.de
[2] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, Karlsruhe, {jonas.lachowitzer,matthias.budde}@disy.net
[3] Bytefabrik.AI GmbH, Haid-und-Neu-Str. 10-14, Karlsruhe, patrick.philipp@bytefabrik.ai

doi:10.18420/inf2022_27

have a large number of parameters and no single visualization for a given dataset is likely to be optimal for *all* users in *all* contexts.

In this work, we introduce the concept of *AIDA-Vis*, a system that generates visualization descriptions that are more aligned with user preferences. To that end, we first pre-train the system via supervised learning on a corpus of labeled examples. Next, pairs of visualizations are generated for a given dataset and shown to a user, who selects one of them to indicate his preferences. This feedback is collected and iteratively incorporated into our system.

2 Problem Statement

The main goal is to create a visualization generator that yields a description v_{desc} for a given table. A table is represented by a dataset D with m rows $R = \{r_i, ..., r_m\}$ and n columns $C = \{c_i, ..., c_n\}$. Usually, human data analysts create a visualization by selecting a subset of columns from C, to display on the y-axis $C_y = \{c_i^y, ..., c_n^y\}$ and another subset, to display on the x-axis $C_x = \{x_i^x, ..., x_n^x\}$. Moreover, analysts select a fitting trace type $t \in T = \{t_i, ..., t_n\}$ (here: scatter plot, line chart, bar chart). Furthermore, data analysts choose a subset l_{sub} from a list of layout properties $L = \{l_i, ..., l_n\}$ (e.g., color of lines in line chart, size of points in a scatter plot, width of bars in a bar chart, etc.)[Hu19]. To automate the process of data visualization, the system should be able to output the description $v_{\text{desc}} = \{C_y, C_x, t, l_{\text{sub}}\}$ for a given dataset D. The description can then be used to render a visualization.

In order to satisfy the requirement that the system should generate visualization descriptions that are optimized for the individual user's preferences, users need to be able to express themselves and the system should be updated to take those preferences into account.

The so-called cold-start problem arises when a system generates mostly arbitrary, useless output, while there is not enough data available yet. To avoid the cold-start problem in this work, the system should generate preliminary visualization descriptions of acceptable quality, without the need for collecting large amounts of user preferences first.

3 Methods

Fig. 1 shows how the components of our approach interact. The core idea of our work is to learn a model first that is able to predict a score for a given visualization description v_{desc}. The score s describes the quality of v_{desc} for a given dataset. This model, then, acts as a *utility function* $u : v_{\text{desc}} \rightarrow \mathbb{R}$ for a second model, the *visualization description generator model* $g : D \rightarrow v_{\text{desc}}$. Now, the model g is learned to output visualization descriptions that maximize the utility function. To prevent the cold-start problem, the model u is pre-trained on labeled examples first. Finally, both models are incrementally improved by (1) collecting human preferences, (2) retraining the model u with the collected preferences, and (3) retraining the visualization description generator model g on the updated utility function. Incrementally improving our model u with human preferences was inspired by [Ch17].

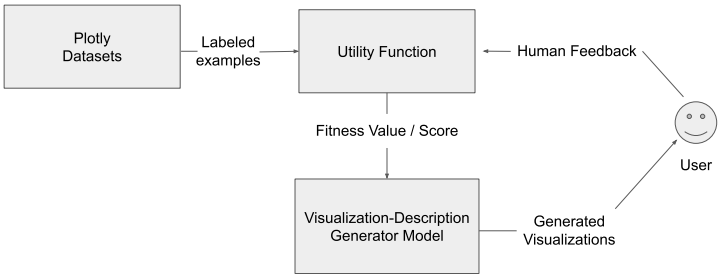


Fig. 1: Components in the learning process. A user provides feedback to a utility model, which in turn approximates the utility for the user. The outcome is then used as a fitness function to train a visualization description generator model, which finally creates visualization recommendations.

3.1 Utility Function & Cold-Start Problem

The utility function is represented by a regression model, which predicts the score for a given visualization description. We use a random forest algorithm to train the model. Our corpus is based on the *Plotly* community feed [Pl], containing human-created pairs of datasets and the corresponding visualization descriptions. We use those pairs as positive examples, for which we set the score to optimal (represented by the scalar 1).

Next, we synthetically generate the non-optimal or negative examples. For a given pair's dataset, an arbitrary visualization description is created. Afterwards, the generated visualization descriptions are compared to the pair's original one. Finally, the scores for the created description are calculated by following some simple hand-crafted rules to quantify the difference. We repeat this procedure for all n datasets. Once we have positive and negative examples, we are able to solve the cold-start problem by pre-training the regression model. The model can already be used in generating acceptable visualizations, which can be improved by collecting human preferences as described later.

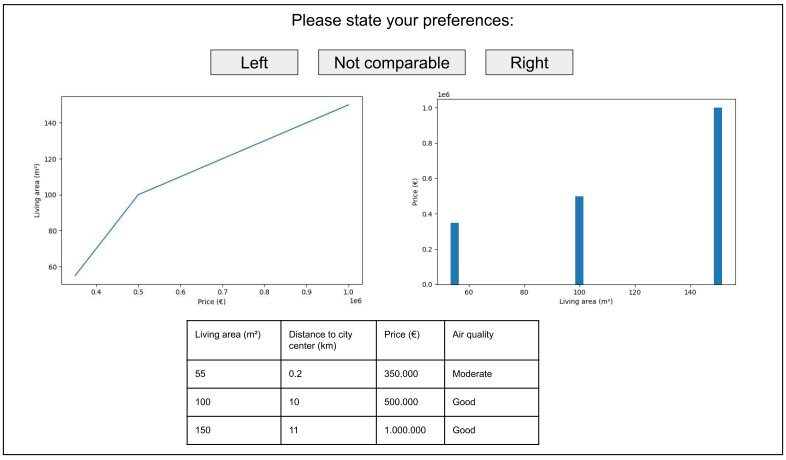
3.2 Training of the Visualization Description Generator Model

The visualization description generator model uses a traditional evolutionary algorithm to optimize the pre-trained utility function u [We15]. The algorithm is initialized with a population p consisting of n randomly generated possible solutions (arbitrary visualization descriptions) $p = \{v^1_{desc}, ..., v^n_{desc}\}$. Those candidate solutions p are assigned a fitness value by the utility function u. According to those fitness values, the candidates are chosen by tournament selection, resulting in a subset $o \in p$. To evolve superior candidate solutions, the selected solutions o are recombined. The recombination is an adaption of the uniform crossover. Thereby, two selected solutions swap either C_x, C_y, t, or l_{sub} with a probability p_{cx}. The resulting solutions are mutated with a small probability p_{mut} by randomly changing C_x, C_y, t, or l_{sub} within the given range by D. Afterwards, the solutions are re-evaluated.

The steps are repeated until the maximum number of iterations is met or a specific fitness threshold is passed. Thereby, each iteration generates candidate visualization descriptions, which then could be used to render the corresponding visualizations.

3.3 Human Feedback

In order to generate customized visualizations for individual users, we added a mechanism for incorporating human preferences. First, the human preferences need to be collected. This can be done by showing two different rendered visualization descriptions for the same dataset to the human via a web interface (see Fig. 2). The user then selects the preferred visualization or indicates that the visualizations are not comparable. The preferences (v^1, v^2, μ) are stored in a database, where v^1 and v^2 are the two different visualization descriptions and μ is a distribution over $\{1, 2\}$ indicating which visualization description the user prefers [Ch17]. The collected data is then used for updating the regression model.



Living area (m²)	Distance to city center (km)	Price (€)	Air quality
55	0.2	350.000	Moderate
100	10	500.000	Good
150	11	1.000.000	Good

Fig. 2: Two rendered visualization descriptions with the corresponding dataset (table) are shown to the user. By clicking on one of the three buttons (*Left*, *Not comparable*, *Right*), the user states his preferences. Afterwards, new visualizations and the corresponding table are shown.

For simplicity, we created the human preferences synthetically in this work. We plan to use real human data in future work.

4 Conclusion & Future Work

We developed and implemented a proof-of-concept of *AIDA-Vis*, a system that automatically generates visualization descriptions for a given dataset. To be able to generate initial, comparable visualization descriptions of acceptable quality, the system is pre-trained on a

corpus based on the *Plotly* community feed [Pl]. The system's output can then be improved and refined by iteratively collecting and incorporating user preferences, which leads to more user-aligned outputs. This is an improvement to current, more rigid approaches, which don't take user preferences into account [DD19; Hu19; Lu18; Mo19].

For future work, we plan to incorporate real human preferences into our *AIDA-Vis* system and compare it to other works like [Hu19]. Moreover, we intend to experiment with various active learning approaches that would allow finding visualization descriptions with the highest information gain. By always showing visualizations with the highest information gain when asking for human feedback, the total needed number of human feedback could be reduced.

Acknowledgments This work was carried out with the support of the German Federal Ministry of Education and Research (BMBF) within the project „AIDA-Vis" (Grant 01IS20097).

References

[Ch17] Christiano, P. F.; Leike, J.; Brown, T. B.; Martic, M.; Legg, S.; Amodei, D.: Deep Reinforcement Learning from Human Preferences. In: Advances in Neural Information Processing Systems 30. Pp. 4299–4307, 2017.

[DD19] Dibia, V.; Demiralp, Ç.: Data2Vis: Automatic Generation of Data Visualizations Using Sequence-to-Sequence Recurrent Neural Networks. IEEE Computer Graphics and Applications 39/5, pp. 33–46, 2019.

[Hu19] Hu, K. Z.; Bakker, M. A.; Li, S.; Kraska, T.; Hidalgo, C. A.: VizML: A Machine Learning Approach to Visualization Recommendation. In: CHI. ACM, 2019.

[Lu18] Luo, Y.; Qin, X.; Tang, N.; Li, G.: DeepEye: Towards Automatic Data Visualization. In: 34th IEEE ICDE 2018. Pp. 101–112, 2018.

[MHS07] Mackinlay, J. D.; Hanrahan, P.; Stolte, C.: Show Me: Automatic Presentation for Visual Analysis. IEEE Trans. Vis. Comput. Graph. 13/6, pp. 1137–1144, 2007, URL: https://doi.org/10.1109/TVCG.2007.70594.

[Mo19] Moritz, D.; Wang, C.; Nelson, G. L.; Lin, H.; Smith, A. M.; Howe, B.; Heer, J.: Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco. IEEE Trans. Vis. Comput. Graph. 25/1, 2019.

[Pl] Plotly-Community: Plotly | Make charts and dashboards online, https://chart-studio.plotly.com/feed/, (Accessed on 07/06/2022).

[We15] Weicker, K.: Evolutionäre Algorithmen. Springer-Verlag, 2015.

[Wo16] Wongsuphasawat, K.; Moritz, D.; Anand, A.; Mackinlay, J. D.; Howe, B.; Heer, J.: Voyager: Exploratory Analysis via Faceted Browsing of Visualization Recommendations. IEEE Trans. Vis. Comput. Graph. 22/1, pp. 649–658, 2016, URL: https://doi.org/10.1109/TVCG.2015.2467191.

Nutzenbasierte Preisgestaltung kooperativer KI-basierter Software in produzierenden Unternehmen – Eine empirische Untersuchung KI-basierter Geschäftsmodelle

Patrick Berger[1], Prof. Dr. Jörg von Garrel[2]

Abstract: Mit Hinblick auf KI als Schlüsseltechnologie befasst sich dieser wissenschaftliche Beitrag, im Rahmen der nutzenbasierten Preisgestaltung, mit der Identifikation von Nutzentreibern auf die Kaufentscheidung einer kooperativen KI (als erklärbare KI), sowie der Analyse der Zahlungsbereitschaft. Neben der ökonomischen Dimension mit Bezug auf Nützlichkeit und Nutzbarkeit des Systems, steht vor allem der (innovative) kooperative Charakter im Fokus. Die Analyse erfolgt mittels einer Choice-based Conjointanalyse (CBC) am Beispiel eines intelligenten Assistenzsystems für Mitarbeitende, dass bei unternehmensinternen Prozessen unterstützt. Hierfür wurden fiktive Kaufangebote erstellt unter denen Entscheidungsträger*innen in produzierenden Unternehmen in Deutschland simulierte Kaufentscheidungen getroffen haben.

Keywords: Kooperative KI, Nutzentreiber, Zahlungsbereitschaft, Choice-based Conjointanalyse, nutzenbasierte Preisgestaltung,

1 Ausgangslage und Problemstellung

Künstliche Intelligenz (KI) zählt als Schlüsseltechnologie zu den Treibern der Digitalisierung und bietet neue Entwicklungspotenziale hinsichtlich Prozess- und/oder Produktinnovationen sowie (Service-)Geschäftsmodellen industrieller Produkte, welche mit intelligenten (digitalen) Services erweitert werden (z.B. Prozesse intelligent überwachen, steuern und regeln, sie flexibler machen, um so Industrie 4.0 auf eine neue Stufe zu heben) [GJS22]. Nichtsdestotrotz ist die Diffusion der KI-Nutzung auf Anwenderseite in deutschen produzierenden Unternehmen verhältnismäßig gering [Ra20]. Begrenzte Ressourcen wie Kapital, Fachexpertise und Weiterbildungsmöglichkeiten für Mitarbeitende hemmen die Integration und Diffusion im produzierenden Sektor [MTG22]. Häufig wird in diesem Kontext auch der Umstand genannt, dass KI-Systeme für Nutzende in einer „Black-Box“ abstrahiert werden. Dies führt dazu, dass von KI-Systemen generierte Ergebnisse nicht mehr oder nur mit unverhältnismäßig hohem Aufwand nachvollziehbar sind und damit als intransparent wahrgenommen werden. [BD17]. Gerade in sensiblen Anwendungsbereichen in der

[1] Hochschule Darmstadt, University of Applied Sciences, wissenschaftlicher Mitarbeiter, Patrick.Berger@h-da.de
[2] Hochschule Darmstadt, University of Applied Sciences, Fachbereich Gesellschaftswissenschaften, Haardtring 100, D - 64295 Darmstadt, joerg.vongarrel@h-da.de
https://orcid.org/0000-0002-3617-1798

doi:10.18420/inf2022_28

Industrie (u.a. bei (strategischen) Entscheidungsunterstützungen) kann die Nachvollziehbarkeit sowie Begründung – und damit die Transparenz – KI-genierter Handlungsvorschläge, ein zentrales Kriterium der Nutzung sein. [GK20]. Dadurch gewinnen bei der Implementierung und dem Betrieb von KI-Systemen im industriellen Kontext Faktoren wie Vertrauen und Akzeptanz an zentraler Bedeutung [JG21]. Auf Anbieterseite werden vor diesem Hintergrund innovative Entwicklungsmethoden wie kooperative KI umgesetzt, um diesem Hemmnis entgegenzuwirken [Ss21]. Wesentliche Merkmale der kooperativen KI sind, dass sich lernende Systeme erklären („Erklärbare KI“), und dass sie sich durch Interaktion mit dem Mitarbeitende an den Menschen anpassen können („Interaktive KI“) [AB18; Ho16]. Aufgrund des innovativen Charakters dieser KI-basierten Softwarelösungen besteht aber wenig Erfahrung darüber, ob die Eigenschaften kooperativer KI-Systeme als Nutzentreiber Kaufentscheidungen fördern. Ziel dieses Artikels ist, relevante Nutzentreiber kooperativer KI-Systeme zu identifizieren sowie deren Einfluss auf die Zahlungsbereitschaft in produzierenden Unternehmen in Deutschland zu analysieren[3]. Mittels einer Choice-based Conjointanalyse (CBC) am Beispiel eines intelligenten Assistenzsystems, erfolgt die Analyse der Nutzenpräferenzen sowie der Zahlungsbereitschaft, mit Bezug auf relevante Eigenschaften kooperativer KI-Systeme für Mitarbeitende in produzierenden Unternehmen in Deutschland[4].

2 Theoretische Grundlagen

Der Begriff KI beschreibt Entwicklungsansätze, die Maschinen kognitive Fähigkeiten beibringen sollen, mit dem Ziel, bestimmte Probleme durch selbständiges Lernen immer besser lösen zu können [BS19]. Auf Anwenderseite stellen aber aktuell die Konfiguration bzw. das Training von Methoden des maschinellen Lernens eine große Herausforderung dar, die oft nur durch spezialisierte Experten bewältigt werden kann. Zum einen gibt es in solchen Anwendungsbereichen strikte Rahmenbedingungen, die von den Systemen eingehalten werden müssen. Zum anderen ist die Auswahl optimaler Modelle eine große Herausforderung. Ein Lösungsansatz besteht in der Entwicklung einer kooperativen KI. Wesentliche Merkmale der kooperativen KI sind, dass es Mitarbeitenden ohne Wissen im Bereich KI möglich sein soll, lernende Systeme zu trainieren, und dass sich diese erklären [Sh19; Ve19] sowie sich durch Interaktion mit dem Mitarbeitenden an den Menschen anpassen können [Na18; Bi18; TK19; Je19]. Hierdurch wird es möglich sein, in Abhängigkeit der Tätigkeit und des Profils eines Mitarbeitenden, lernende Systeme dynamisch entstehen zu lassen, die ihn nicht nur optimal in seiner Tätigkeit unterstützen, sondern auch bessere Ergebnisse in Bezug auf Vertrauen und Akzeptanz generieren [Hi19; MTB19].

[3] Die Untersuchung wurde im Rahmen des KompAKI (https://kompaki.de/) mit Fokus auf produzierenden Unternehmen in Deutschland durchgeführt.

[4] Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 02L19C157 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren

Die zunehmende Automatisierung und (individuelle) Adaptivität des Wertversprechens durch KI, führt aber auch auf Anbieterseite zu Herausforderungen. Klassische Preisermittlungsverfahren stoßen u.a. aufgrund nicht vorhandener Grenzkosten von Daten und Algorithmen an ihre Grenzen [KO18]. Vergleichsstudien von Liozu und Hinterhuber (2012) verdeutlichen, dass Unternehmen mit einem „Value-Based-Pricing" Modell im Gegensatz zu kosten- oder wettbewerbsbasierten Preissetzungsmodellen eine stärkere Preisorganisation aufzeigen, was bedeutet, dass sie sowohl durch das Preiscontrolling als auch durch einen wiederkehrenden Review eine bessere Preisfindung entwickeln können. Die Basis bildet bei den (kunden-)nutzenbasierten Preismodellen eine stärkere systematische Methodik, im Gegensatz zu kostenbasierten Preismodellen, bei denen eher die individuellen Erfahrungswerte und persönlichen Einschätzungen als Grundlage der Preisbildung dienen [LH12].

Die Grundannahme (kunden-)nutzenbasierter Preismodelle basiert darauf, dass Kunden durch Produkt- und Dienstleistungsangebote einen Leistungsvorteil wahrnehmen und anhand dieser Bewertung eine definierte Bereitschaft haben, einen bestimmten Preis zu zahlen – die sogenannte Zahlungsbereitschaft. Diese Zahlungsbereitschaft muss ermittelt werden, um eine optimale Preisbildung vorzunehmen. Dabei gilt es, verschiedene Aspekte des Wertes aus Kundensicht in den Fokus zu rücken [KN91]. Der maximale Preis, den ein Konsument bereit ist für ein Produkt oder eine Dienstleistung zu zahlen (Zahlungsbereitschaft), korrespondiert unmittelbar mit dem wahrgenommenen Kundennutzen, welcher dieses Produkt für den Konsumenten hat [KN91]. Die Zahlungsbereitschaft kann somit als monetärer Ausdruck des wahrgenommenen Kundennutzens aufgefasst werden [KN91].

Eine umfangreiche Literaturrecherche relevanter Nutzentreiber kooperativer KI-Systeme im B2B-Kontext hat unterschiedliche Eigenschaften in den Kategorien *allg. Systemeigenschaften (KI)*; *Erklärbarkeit*; *Interaktivität*; *Autonomie*; *Transparenz*; *Datenverarbeitung*; *Monetäre Dimension (KI-basierter) Software* identifiziert [Ho19; JN18; SSD17; SP21; Zi21; KO18; CF21; Ke21; MZH19; Ra21; Sa19; KWS21; In21].

Kategorie	Eigenschaften		
Allgemeine Systemeigenschaften KI-basierter Software	Trainingshäufigkeit	Individualisierungsgrad (KI-Modell)	Lernverfahren
	Trainingsaufwand (=Zeit + Rechenleistung)	Algorithmus-Performance	Evaluation/ Übersicht der Metriken
	Latenz/ Rechenzeit pro Anfrage	Systemaktivierung	Systemverfügbarkeit
	Funktionalitäten	Wertversprechen KI-basierter Software	Simulationsfunktion
Erklärbarkeit	Erklärungsspezifikation des (KI-) Modells	Erklärungsmethode des Algorithmus (XAI Method)	
Interaktivität	Interaktion des Nutzers mit dem System	Grad der Interaktivität	Grad der Gamifizierung
	Benutzerfreundlichkeit		
Autonomie	Algorithmus-Autonomität/ Grad der Proaktivität	Grad der Selbstüberwachung	
Transparenz	Umfang der Datenerfassung	Darstellung der Datennutzung (ggü. Mitarbeitenden)	Datenzugriff (intern)
Datenverarbeitung	Datenverarbeitung (Lokal/ Cloud)	Datenübertragung bei Cloudlösungen	Anbieterkapazität bei Cloudlösungen
Monetäre Dimension (KI-basierter) Software	Integrationsaufwand	Implementierungskosten	Erlösmodelle & Preisstrategien

Abb. 1: Kaufrelevante Eigenschaften (kooperativer) KI-basierter Software

2.1 Choice-based Conjointanalyse (CBC)

Die methodische Durchführung erfolgte mittels einer Choice-Based Conjointanalyse nach Louviere und Woodworth [LW83]. Sie hat das Ziel: *„(…), Kaufentscheidungen von Konsumenten über eine dekompositionelle Schätzung der Bewertung von Produktmerkmalen zu erklären.“* [Ba21]. Sie unterscheidet sich methodisch von der Grundlage der klassischen Conjointanalyse (TCA). Im Vergleich zur TCA werden bei der CBC keine ordinalen oder metrischen Präferenzurteile über Eigenschaften getroffen, sondern (fiktiv) diskrete Kaufentscheidungen vorliegender Produktprofile (Choice-Tasks/ Stimuli) analysiert [Ba21]. Als Voraussetzung gilt die Wahl der Alternative, die dem Konsumenten den relativ höchsten Nutzenwert (=Nettonutzen) bringt. Dabei wird von den beobachteten Auswahlentscheidungen auf Nutzenvorstellungen der Konsumenten durch einzelne Produktmerkmale geschlossen. Aus theoretischer Sicht wird daher der CBC eine höhere Anreizkompatibilität bezüglich der Offenbarung der tatsächlichen Kaufabsicht und Zahlungsbereitschaft zugesprochen [Da12]. Die eigentliche Kaufentscheidung wird letztlich auf Grund der Leistung eines Objekts getroffen. Die Ermittlung der Nutzen- und Teilnutzenwerte der Eigenschaften und Ausprägungen erfolgt über die hierarchische bayesianische Schätzung (Maximum-Likelihood-Methode). Dabei beschreibt die Likelihood-Funktion (aggregierte Wahrscheinlichkeitsfunktion) den Zusammenhang zwischen objektiven Eigenschafts-ausprägungen (verschiedener Leistungsbündel) und dem subjektiven Nutzen [Ge06].

Für die Gestaltung geeigneter Choice-Sets sind folgende Gütekriterien einzuhalten [Da12]:

1. *Empirische Unabhängigkeit* → kein empirischer Wirkungszusammenhang
2. *Präferenzrelevante Eigenschaften der Kaufentscheidung*
3. *Variation der Ausprägungen* → Variation der Ausprägungen je Eigenschaft
4. *Begrenzung der Eigenschaften* → Komplexitätsreduktion
5. *Kompensatorische Merkmalsbeziehungen* → Bspw. hohe Qualität → hoher Preis

Um diesen Gütekriterien Ziel gerecht zu werden, erfolgte eine Verdichtung der identifizierten Eigenschaften. Hierzu sind mittels Expertengespräche auf Anbieter- als auch Entwicklerseite sechs kaufrelevante Eigenschaften einer kooperativen KI-basierten Software für das weitere Vorgehen isoliert und hinsichtlich ihrer Ausprägungen an das Untersuchungsobjekt operationalisiert worden (Abb.2). Die gewählten Eigenschaften beziehen sich zum einen auf die im B2B-Kontext zentrale ökonomische Dimension [Th17] sowie den wesentlichen Eigenschaften der Kooperativität. Die Bewertungsgrundlage der ökonomischen Dimension bezieht sich neben dem Preis auf die Nützlichkeit und Nutzbarkeit eines kooperativen KI-basierten Systems [Th17]. Die Nützlichkeitsbewertung zielt auf die im Performance-Controlling relevanten Indikatoren der Effektivität und Effizienz ab. Die Nutzbarkeitsbewertung bezieht sich dabei zum einen auf die Prozesskompatibilität der Software und zum anderen auf die Aufwandskomplexität der nutzenorientierten Entwicklung. Zur realitätsgetreuen Simulation verteilt sich der Preis in Abhängigkeit der verursachten Aufwände und Kosten prozentual auf die Eigenschaftsausprägungen im Sinne einer bedingungsabhängigen Preissetzung[5] auf Basis eines Erlösmodells[6] sowie einer Preisstrategie[7]. Prinzipiell gilt die Preispolitik im B2B als ein sensibles Thema und ist öffentlich nur schwer oder limitiert zugänglich. Meist gibt es keine endgültige zielgruppenspezifische Preispolitik, sondern diese wird zwischen den Organisationen abhängig vom Use Case verhandelt [JN18]. Trotz der limitierten Verfügbarkeit von öffentlich zugänglichen Preisen sind drei Produkte als Orientierungshilfe für den Use Case identifiziert worden[8]. Ausgehend von diesen Marktpreisen ergibt sich eine Preispanne von 300€ - 500€ mit

[5] Die bedingungsabhängige Preissetzung ermöglicht eine realitätsnähere Bestimmung der Zahlungsbereitschaft einzelner Eigenschaftsausprägungen

[6] Das Erlösmodell stellt eine Kombination aus „pay-per-use“ (pro 1000 Anfragen) und „pay-per-function“ dar [Ho19; Zi21; Ko18; JN18]. Die Preisvariationen im „pay-per-function“ ergeben sich abhängig von anfallenden Aufwänden und Kosten je Funktionalität. Es wird sowohl das Verhältnis der jeweiligen Eigenschaftsausprägungen zueinander, als auch der Eigenschaften untereinander berücksichtigt. Beispielsweise würde die Entwicklung und das trainieren von Modellen erheblich höhere Aufwände und Kosten verursachen als eine transparente Darstellung/Wiedergabe der Erklärungsmethoden. Dabei würde das Entwickeln eines individuellen Modells höhere Programmier- und Trainingsaufwände bedeuten und damit letztlich mehr Kosten verursachen, als ein herkömmliches Standardmodell.

[7] Die Preisstrategie basiert auf dem „market-based-pricing“ [JN18; In21]

[8] Hierbei handelt es sich um: „Conversational AI“ (SAP-Store); „Azure Bot“ (Azure-Microsoft-Store); „Amazon Lex“ (AWS-Marketplace).

einem durchschnittlichen Markpreis von ((300×2) + 500) / 3=366,67 → 370€ pro 1000 Anfragen. Die dargestellten Eigenschaftsausprägungen tragen Preisanteile, die in Summe am Ende, Produktausführungen in der ermittelten Preisspanne ergeben[9]. Die Preisfindung einer fertigen Produktausführung basiert auf einem Basispreis von 25€ pro 1000 Anfragen, plus den Teilkosten/Preisanteilen einzelner funktioneller Bestandteile der Software.

Eigenschaft	Ausprägung 1	Ausprägung 2	Ausprägung 3	Indikator
Interaktion	Verbal **(25€)**	Schriftlich/ Icon-basiert **(25€)**	Verbal + Schriftlich/ Icon-basiert **(50€)**	*Interaktivität*
Grad der Interaktivität	Keine Nutzer-anpassung (statisch) **(25€)**	Anpassung an allg. Mitarbeiterdomäne **(50€)**	Personalisierte Nutzeranpassung **(100€)**	*Interaktivität*
Individualisierungsgrad (KI-Modell)	Standardmodell **(100€)**	Branchenmodell + Unternehmens-spezifische Trainingsdaten **(150€)**	Individuelles Modell + Unternehmens-spezifische Trainingsdaten **(225€)**	*Effektivität; Prozess-kompatibilität; Aufwands-komplexität*
Transparenz des KI-Modells	Modell erklärt sich nicht **(25€)**	Grundlegend statistische Erklärungsmethoden **(50€)**	Schlussfolgernde Erklärungen des Algorithmus **(100€)**	*Erklärbarkeit/ Transparenz*
Algorithmus-Performance	Gering (89%-90%) **(25€)**	Moderat (94%-95%) **(25€)**	Hoch (99%) **(75€)**	*Effektivität; Effizienz*
Preis (pro 1000 Anfragen)	= Basispreis + Summe der Teilkosten je Eigenschaftsausprägung			

Abb. 2: Bedingungsabhängige Preisgestaltung eines KI-basierten Assistenzsystems

Konkret umgesetzt wurde die CBC anhand eines intelligenten Assistenzsystems im Sinne eines unternehmensinternen Chatbots, das Mitarbeitende bei internen Prozessen unterstützen kann. Dieses KI-basierte System kann mit verschiedenen unternehmens-internen Systemen (z.B. ERP-, IoT-Systemen, Predicitve Analytics, Projektmanagement-Tools, etc.) über API´s verknüpft werden und dadurch Workflows optimieren. Es ist in der Lage, mit Mitarbeitenden zu interagieren und bei Tätigkeiten zu unterstützen die auf abteilungsübergreifenden Informationen basieren, wie zum Beispiel:

- „Schicke Mail X an Mitarbeitende, die an Prozess Y beteiligt sind."
- „Mach einen Termin mit XY für *Datum*/ *Uhrzeit*."

[9] Wenige Produktausführung befindet sich außerhalb der identifizierten Preisspanne mit einem Mindestpreis von 225€ pro 1000 Anfragen und einem Höchstpreis von 575€ pro 1000 Anfragen.

- „Zeig mir das Live-Monitoring von Prozess C."

Auf Basis von Abb. 2 (3^5= 243 Kombinationsmöglichkeiten) wurde ein reduzierter orthogonaler Versuchsplan mit einer benötigten Stichprobengröße von *n*= 100 erstellt. Zur Generierung eines Datensatzes mit bewertungstauglichen Probanden ist die Zielgruppe *„Entscheidungsträger*innen in produzierenden Unternehmen in Deutschland"* identifiziert worden. 162 Personen eines der Zielgruppe entsprechenden Panels des Dienstleisters *KANTAR Group* [Ka22] haben an der Umfrage teilgenommen. In 10 fiktiven Kaufentscheidungssituationen konnten die Probanden zwischen je 3 Angeboten auswählen.

	Angebot 1	Angebot 2	Angebot 3
Interaktion	Schriftlich (bzw. Icon-basiert)	Schriftlich (bzw. Icon-basiert)	Schriftlich (bzw. Icon-basiert)
Interaktivität (Anpassung der Interaktion an den Menschen)	Personalisierte Nutzeranpassung (hoch)	Anpassung an allg. Mitarbeiterdomäne (mittel)	Keine Nutzeranpassung (statisch)
Individualisierungsgrad (der Software)	Individuelles Modell + unternehmensspezifische Trainingsdaten (hoch)	Branchenmodell + unternehmensspezifische Trainingsdaten (mittel)	Standardmodell (gering)
Transparenz des KI-Modells	Grundlegend statistische Erklärungsmethoden (gering)	Schlussfolgernde Erklärungen des Algorithmus (hoch)	Grundlegend statistische Erklärungsmethoden (gering)
Algorithmus-Performance (Zuverlässigkeit)	Moderat (94-95%)	Hoch (99%)	Moderat (94-95%)
Preis (pro 1000 Anfragen)	€450	€425	€250
	○	○	○

Würden Sie die ausgewählte Option kaufen?

Ja

Nein

Abb. 3: Beispiel einer simulierten Kaufentscheidungssituation

3 Ergebnisse

Die Stichprobe setzt sich aus 59,0% (*n*= 95) KMU und 41,0% (*n*= 66) Großunternehmen zusammen. Diese umfassen nach NACE-Codes folgende Branchen (siehe Abb. 4). Insgesamt haben Probanden mit verschiedenen Unternehmenspositionen an der Umfrage teilgenommen (Geschäftsführer= 14,4%/*n*=23; Angestellte mit Führungsverantwortung= 70%/*n*=112; Angestellte ohne Führungsverantwortung= 15,6%/*n*=25). Unter diesen gaben dreiviertel (74,7%) an, am Beschaffungsprozess von Software im Unternehmen beteiligt zu sein. Wie der grafischen Darstellung der Branchenverteilung *(n*= 217) zu entnehmen ist, sind gewisse Unternehmen branchenübergreifend tätig.

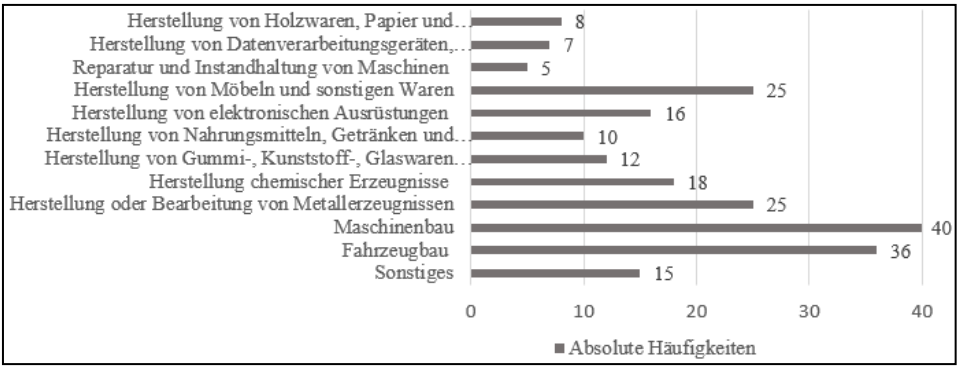


Abb. 4: Repräsentierte Branchen nach NACE-Codes

Insgesamt haben circa dreiviertel (72,84%) der Befragten in 7 bis 10 der Kaufentscheidungssituationen eines der dargestellten Angebote zum Kauf ausgewählt. Knapp die Hälfte (49,3%) hat dies in jeder Kaufentscheidungssituation getan. Ausschließlich 7,4% (*n*= 12) haben in keiner der simulierten Kaufentscheidungssituationen einen Kauf getätigt (siehe Abb. 5).

Die Auswertung der Ergebnisse[10] in Abb. 7 zeigt die Teilnutzenwerte[11] in Abhängigkeit der Kosten (Preisanteile) und beschreibt somit die Zahlungsbereitschaft einzelner Eigenschaftsausprägungen. Insgesamt werden 8 der 15 Eigenschaftsausprägungen ein positiver Nettonutzen in Abhängigkeit des Preisanteils zugesprochen.

[10] Die technische Umsetzung sowie Durchführung und Auswertung erfolgte mit dem Umfrage- und Statistiktool *qualtrics*^{XM}.

[11] Die Präferenz einzelner Eigenschaftsausprägungen stellt sich durch die geschätzten Teilnutzenwerte (=Nettonutzen→ NN) dar. Diese beschreiben mit welcher Wahrscheinlichkeit (in Abhängigkeit des Preisanteils), die Eigenschaftsausprägung eine Produktkombination hinsichtlich einer Kaufabsicht (wahrscheinlich) verbessern oder verschlechtern würde.

Die Eigenschaftsausprägungen mit dem höchsten Teilnutzenwert je Eigenschaft in Kombination, entsprechen der optimalen Produktkombination mit dem höchsten erwarteten Nettonutzen zur maximalen Zahlungsbereitschaft (Abb. 6). Es repräsentiert letztlich die Produktentwicklung mit der höchsten Kaufwahrscheinlichkeit (Abb. 7 rote Kästchen).

$$\mathbf{NN_A}^{12} > \mathbf{NN_B}^{13} + \mathbf{U}^{14}$$

Betrachtet man Eigenschaftsausprägungen mit einem positiven Teilnutzenwert, weist das „*Branchenmodell + unternehmensspezifisches Training*“ in jeder möglichen Produktkombination den größten Preisanteil auf. Gleichzeitig wird dieser Eigenschaftsausprägung in Relation aber ein höherer Teilnutzenwert zugeschrieben. Es hebt sich hervor, dass die Probanden gegenüber dem erwarteten Nutzen dieser Eigenschaftsausprägung (150€ pro 1000 Anfragen) die höchste Zahlungsbereitschaft zeigen.

Positive Kaufentscheidungen	Häufigkeiten
0	*n= 12*
1	*n= 3*
2	*n= 3*
3	*n= 4*
4	*n= 4*
5	*n= 5*
6	*n= 13*
7	*n= 17*
8	*n= 11*
9	*n= 10*
10	*n= 80*

Abb. 5: Anzahl positiver Kaufentscheidungen

Optimale Produktkombination (*n*= 162)
Verbal + Schriftlich (bzw. Icon-basiert)
Anpassung an allg. Mitarbeiterdomäne (mittel)
Branchenmodell + unternehmensspezifisches Training (mittel)
Schlussfolgernde Erklärung des Algorithmus (hoch)
Hoch (99%)
450€ (pro 1000 Anfragen)
Effektstärke: Stark

Abb. 6: Optimale Produktkombination

[12] $\mathbf{NN_A}$= Nettonutzen von Angebot A
[13] $\mathbf{NN_B}$= Nettonutzen von Angebot B
[14] **U**= Status quo (aktueller Nettonutzen)



Abb. 7: Darstellung der geschätzten Teilnutzenwerte in Abhängigkeit des Preisanteils (*n*= 162).

Zusätzlich kann Abb. 7 entnommen werden, dass die Probanden vor allem ein interaktives System bevorzugen, dass sich in einem gewissen Rahmen (Mitarbeiterdomäne) an die Nutzenden anpasst und auf verschiedene Arten Interaktionsmöglichkeiten (verbal + schriftlich) bietet. Den stärksten Nutzen nehmen die Probanden für eine hohe Zuverlässigkeit (Algorithmus-Performance) des Systems wahr. Dies lässt sich u.a. durch die Erlösmodellkomponente „pay-per-use“ nachvollziehen, da eine geringe Performance (mehr falsch-klassifizierte Anfragen) langfristig zu höheren Nutzungskosten führen.

Darüber hinaus verdeutlicht Abb. 7, dass die Probanden eine transparente KI bevorzugen. Während „*Grundlegend statistische Erklärungsmethoden*“ und

„*Schlussfolgernde Erklärungen des Algorithmus*“ einen positiven Teilnutzenwert hervorbringen, weist die Eigenschaftsausprägung „*Modell erklärt sich nicht*“ einen negativen auf. Genauer nehmen die Probanden bezüglich der Transparenz den stärksten Nutzen für ein sich selbsterklärendes System wahr und sind bereit dafür 100€ pro 1000 Anfragen zu bezahlen.

4 Diskussion, Fazit und Ausblick

Zur Diskussion stehen aus Sicht der Autoren hinsichtlich der Validität und Aussagekraft der Ergebnisse, eine fehlende Pauschalisierung der Ergebnisse für andere KI-basierte Softwarelösungen. Des Weiteren umfasst die CBC-Analyse (auf Grund der notwendigen Komplexitätsreduktion) nur eine sehr geringe Anzahl an inkludierten Eigenschaften, die nicht alle kaufrelevanten Eigenschaften abdeckt. Ebenfalls ist ein hypothetischer Bias denkbar, da mit den Kaufentscheidungen keine realen Transaktionen verbunden waren. Die Reliabilität ist im Rahmen der Preisgestaltung zu bemängeln, da die Preisfindung im Kontext von B2B meist Use Case abhängig und schwer verallgemeinerbar ist. Die hohe Anzahl an positiven Kaufentscheidungen lässt darauf schließen, dass die erstellten Angebote zu einem zu geringen Preis angeboten wurden oder eine sehr hohe Nachfrage für solche KI-basierten Softwarelösungen am Markt besteht. Es könnte aber auch durch die vertretene Anzahl an Großunternehmen erklärt werden, die über mehr finanzielle Mittel verfügen und daher die Kosten als geringeres Risiko gegenüber dem Nutzen einschätzen.

KI-basierte Softwarelösungen zählen zu den Schlüsseltechnologien unserer Zeit und bieten innovative Lösungsansätze für komplexe Problemstellungen. Trotz wissenschaftlicher Erkenntnisse aus Theorie und Praxis weist die Diffusion in deutschen Unternehmen weiterhin großes Potential auf. Die zukünftig steigende Zahl von KI-Einsätzen erfordert Mitarbeitende, die solche Systeme trainieren und kontrollieren können. Der Arbeitsmarkt für diese Fachkräfte ist äußerst angespannt, (insbesondere im Rhein-Main-Gebiet,) sodass es für (kleine) produzierende Gewerbe sehr schwierig ist diese zu akquirieren. Für den praxisorientierten und menschengerechten Einsatz in der breiten Arbeitswelt ist es aber essentiell, Formen von KI zu entwickeln, die von Nicht-Experten bedient und verstanden werden können. Laut empirischen Befunden scheint vor allem das abstrahieren von KI-Systemen in „Black-Boxen“, auf Grund intransparenter Entscheidungsfindung der Algorithmen, den Einsatz zu hemmen. Ein möglicher Lösungsansatz zur Kompensation stellt die kooperative als erklärbare KI dar, die zusätzlich das Vertrauen der Menschen in KI-basierte Systeme (in der breiten Masse) erhöhen soll und dadurch die Akzeptanz für den KI-Einsatz in deutschen Unternehmen stärkt. Die Auswertung der CBC-Analyse verdeutlicht, dass Entscheider*innen in produzierenden Unternehmen neben einem hohen Interaktivitätsgrad sowie einem hohen Zuverlässigkeitsniveau, der Transparenz im Sinne des Erklärungsgehaltes einen hohen Nutzen beimessen.

Die dargestellten Ergebnisse bilden den Startpunkt für weitere Analysen. So ist es das Ziel Faktoren wie „Erfahrungen mit KI-Systemen“ aber auch „Unternehmensgröße“ als unabhängige Variablen in die weiterführenden Analysen einzubringen und damit weitere nutzenbestimmende Faktoren zu identifizieren. Zusätzlich sind weitere CBC-Analysen mit weiteren, beispielhaften KI-Systemen geplant, die auch hier weiteren Aufschluss über eine nutzenbasierte Preisgestaltung kooperativer KI-basierter Software geben werden.

5 Literaturverzeichnis

[AB18] Adadi, A.; Berrada, M. (2018): Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI), IEEE Access, Volume 6, 52138-52160 [Ba21] Balderjahn I., Hedergott D., Appenfeller D., Peyer M. (2021) Choice-Based Conjointanalyse. In: Baier D., Brusch M. (eds) Conjointanalyse. Springer Gabler, Berlin, Heidelberg.

[Bi18] Binnig, Carsten; Kersting, Kristian; Molina, Alejandro; Zgraggen, Emanuel (2018): DeepVizdom: Deep Inter-active Data Exploration. In Proceedings of the Inaugural Systems and Machine Learning Conference (SysML).

[BD17] Bitkom e.V., DFKI (2017): Entscheidungsunterstützung mit Künstlicher Intelligenz. Wirtschaftliche Bedeutung, gesellschaftliche Herausforderungen, menschliche Verantwortung, Bundesverband Informationswirtschaft, Telekommunikation und neue Medien e.V., Berlin.

[BS19] Buxmann, P., Schmidt, H. (2019): Grundlagen der künstlichen Intelligenz und des maschinellen Lernens. In künstlicher Intelligenz (pp. 3-19). Springer Gabler, Berlin, Heidelberg.

[CF21] Classen, Moritz; Friedli, Thomas (2021): ‘It’s not like we can charge for everything’: revenue models to capture value from smart services in Pacific Asia. In: Asia Pacific Business Review 27 (3), S. 405–430.

[Da12] Dauner, Jörg (2012): Zahlungsbereitschaft für Remote Services. Wiesbaden: Springer Fachmedien Wiesbaden.

[GJS22] Garrel, Jörg von, Jahn, Carlos and Schröter, Dominik (2022): "Der Einsatz Künstlicher Intelligenz in produzierenden Unternehmen: Eine Morphologie industrieller, KI-basierter Arbeitssysteme" Zeitschrift für wirtschaftlichen Fabrikbetrieb, vol. 117, no. 5, 2022, pp. 338-343

[JG21] Jung, Maria; Garrel, Jörg von (2021): Mitarbeiterfreundliche Implementierung von KI-Systemen im Hinblick auf Akzeptanz und Vertrauen. Erarbeitung eines Forschungsmodells auf Basis einer qualitativen Analyse. In: Journal for technology assessment in theory and practice, S. 37–43.

[Ge06] Gensler, S. (2006): Ermittlung von Präferenzen für Produkteigenschaften mit Hilfe der Choice-Based Conjoint Analyse, Teil I. Wirtschaftswissenschaftliches Studium, 35(5), 254.

[GK20] Grünwald, Reinhardt; Kehl, Christoph (2020): Autonome Waffensysteme. In: TAB-Arbeitsbericht Nr. 187 (Hg): Büro für Technikfolgen-Abschätzung beim deutschen Bundestag, Berlin.

[Hi19] Hilbert, Marc; Neukart, Florian; Ringlstetter; Seidel, Christian; Sichler, Barbara (2019): KI-Innovation über das autonome Fahren hinaus. In Buxmann, Peter; Schmidt, Holger (Hg.) (2018): Künstliche Intelligenz. Mit Algorithmen zum wirtschaftlichen Erfolg. Springer Gabler, Berlin.

[Ho19] Hoisl, Barbara (2019): „Software Is Eating the World“ – Software-Märkte ticken anders. In: Barbara Hoisl (Hg.): Produkte digital-first denken. Wiesbaden: Springer Fachmedien Wiesbaden, S. 17–59.

[Ho16] Holzinger, A. (2016): Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Informatics, 3(2), 119-131.

[In21] In, Lee (2021): Pricing and Profit Management Models for SaaS Providers and IaaS Providers. In: JTAER 16 (4), S. 859–873.

[JN18] Jefimova, Zenja; Nabseth, Sofie (2018): A Pricing Model for AIaaS. An analysis of a new AI personalization product within the edtech space. Masterarbeit. KTH ROYAL INSTITUTE OF TECHNOLOGY, Stockholm. KTH Industrial Engineering and Management.

[Je19] Jentzsch, Sophie; Schramowski, Patrick; Rothkopf, Constantin; Kersting, Kristian. (2019): Semantics Derived Automatically from Language Corpora Contain Human-like Moral Choices. In Proceedings of the 2nd AAAI/ACM Conference on AI, Ethics, and Society (AIES).

[KN91] KALISH, S., NELSON, P., (1991): A Comparison of Ranking, Rating, and Reservation Price Measurement in Conjoint Analysis, in: Marketing Letters, S. 327–335.

[Ka22] KANTAR Group.KANTAR.com.07.05.2022

[Ke21] Kett H., Evcenko D., Falkner J., Frings S., Neuhüttler J. (2021): Künstliche Intelligenz als Verände-rungstreiber für Geschäftsmodelle. In: Bruhn M., Hadwich K. (eds) Künstliche Intelligenz im Dienst-leistungsmanagement. Forum Dienstleistungs-management. Springer Gabler, Wiesbaden.

[KO18] Koschate-Fischer, Nicole; Ochmann, Jessica (2018): Erlösmodelle für datenzentrische Dienstleistun-gen. In: Manfred Bruhn und Karsten Hadwich (Hg.): Service Business Development. Wiesbaden: Springer Fachmedien Wiesbaden, S. 169–185.

[KWS21] König, Pascal and Wurster, Stefan and Siewert, Markus B., (2021): Consumers Are Willing to Pay a Price for Explainable, But Not for Green AI. Available at SSRN: https://ssrn.com/abstract=3890774 or http://dx.doi.org/10.2139/ssrn.3890774

[LH12] Liozu, S.M., Hinterhuber, A. (2012): "Industrial product pricing: a value-based approach", Journal of Business Strategy, Vol. 33 No. 4, pp. 28-39

[LW83] Louviere, J.J. und G. Woodworth (1983): Design and Analysis of Simulated Consumer Choice or Al-location Experiments: An Approach Based on Aggregate Data. In: Journal of Marketing Research 20.4, S. 350–367.

[LM19] Lundborg, Martin; Märkel, Christian (2019): Künstliche Intelligenz im Mittelstand – Relevanz, Anwendungen, Transfer (Hg): Begleitforschung Mittelstand-Digital WIK GmbH, Bad Honnef.

[MTG22] Merkel-Kiss, Mirjam & Thomas, Simone & Garrel, Jörg. (2022): Systematische Literaturanalyse zum KI-Einsatz und KI-basierten Geschäftsmodellen in produzierenden kleinen und mittleren Unternehmen. Frühjahrskongress 2022 GfA: Technologie und Bildung in hybriden Arbeitswelten. Magdeburg.

[MTB19] Mesbah, N.; Tauchert, C.; Olt, C. M. Buxmann, P.(2019): Promoting Trust in AI-based Expert Systems. In: Americas Conference on Information Systems - AMCIS 2019, Cancun, Mexico, 15. - 17.08.2019.

[MZH19] Mihale-Wilson, A.C., Zibuschka, J., Hinz O. (2019): User preferences and willingness to pay for in-vehicle assistance. Electron Markets 29, 37–53 (2019).

[Na18] Natarajan, Sriraam; Odom, Phillip; Khot, Tushar; Kersting, Kristian; Shavlik, Jude. (2018): Human-in-the-loop Learning for Probabilistic Programming. In Proceedings of the Inaugural International Conference on Probabilistic Programming.

[Ra21] Ramon, Yanou; Vermeire, Tom; Toubia, Olivier; Martens, David; Evgeniou, Theodoros (2021): Understanding Consumer Preferences for Explanations Generated by XAI Algorithms.

[Sa19] Salah, Khaled; Rehman, M. Habib Ur; Nizamuddin, Nishara; Al-Fuqaha, Ala (2019): Blockchain for AI: Review and Open Research Challenges. In: IEEE Access 7, S. 10127–10149.

[SP21] Schulz, Esther; Pflaum, Alexander (2021): Erfolgsfaktoren bei der Entwicklung datengetriebener Geschäftsmodelle. In: Kai Altenfelder, Dieter Schönfeld und Wolfgang Krenkler (Hg.): Services Management und digitale Transformation. Wiesbaden: Springer Fachmedien Wiesbaden, S. 47–67.

[SSD17] Schüritz, Ronny; Seebacher, Stefan; Dorner, Rebecca (2017): Capturing Value from Data: Revenue Models for Data-Driven Services. In: 50th Hawaii International Conference on System Sciences.

[Sh19] Shang, Zeyuan; Zgraggen, Emanuel; Buratti, Benedetto; Chung, Yeounoh; Eichmann, Philipp; Binnig, Carsten; Upfal, Eli; Kraska, Tim. (2019): Democratizing Data Science through Interactive Curation of ML Pipe-lines. SIGMOD 2019.

[Ss21] Striegel, Serena Simone (2021): Perceived quality in the automotive industry: A Neural Network Based Assessment of Split-lines. Dissertation. Technische Universität Berlin. Berlin

[TK19] Teso, Stefano; Kersting, Kristian (2019): Explanatory Interactive Machine Learning. In Proceedings of the 2nd AAAI/ACM Conference on AI, Ethics, and Society (AIES).

[Th17] Thim, Christof (2017): Technologieakzeptanz in Organisationen. Ein Simulationsansatz. Dissertation. Wirtschafts- und Sozialwissenschaftlichen Fakultät. Universität Potsdam, Potsdam.

[Ve19] Vergari, Antonio; Molina, Alejandro; Peharz, Robert; Ghahramani, Zoubin; Kersting, Kristian; Valera, Isabel. (2019): Automatic Bayesian Density Analysis. In Proceedings

of the AAAI Conference on Artificial Intelligence (AAAI).

[We19] Welsch A., Eitle V., Buxmann P. (2019): Entscheidungsunterstützung im Kundenbeziehungszyklus durch Maschinelle Lernverfahren. In: Meinhardt S., Pflaum A. (eds) Digitale Geschäftsmodelle – Band 2. Edition HMD. Springer Vieweg, Wiesbaden.

[Zi21] Zillner, Sonja (2021): Business Models and Ecosystem for Big Data. In: Edward Curry, Andreas Metz-ger, Sonja Zillner, Jean-Christophe Pazzaglia und Ana García Robles (Hg.): The Elements of Big Data Value. Cham: Springer International Publishing, S. 269–288.

Nachhaltige Wertschöpfungssysteme (NaWerSys) II

2. Workshop: Nachhaltige Wertschöpfungssysteme (NaWerSys)

Thorsten Schoormann [1], Friedemann Kammler [2], Paul Christoph Gembarski [3] und Simon Hagen[4]

1 Einführung und Motivation

In der Entwicklung neuer digitaler Produkte und Dienstleistungen erfolgt gegenwärtig ein kollektives Umdenken hin zu mehr Nachhaltigkeit und Verantwortung [Di20, To20, Lo21, Sc21]. Dies betrifft den gesamten Lebenszyklus (digitaler) Artefakte von der Idee und Gestaltung bis zum Betrieb, zur Nachnutzung und zum Recycling. Informations- und Kommunikationstechnologie nimmt dabei nicht nur die traditionelle Rolle als Intermediär zwischen den anbietenden Unternehmen und dem Kundensegment ein. Sie ist vor allem auch Enabler vernetzter, produktübergreifender Funktionen [Ha19] bis hin zur Autonomisierung ganzer Prozessabläufe, in denen etwa Umweltveränderungen sensorisch erfasst [Lo11], eingeordnet und Handlungsmöglichkeiten abgeleitet werden.

Für die Implementierung solcher übergreifenden Funktionen, bündeln unterschiedliche Unternehmen oft ihre datengetriebenen Produkte und Dienstleistungen und konstituieren Wertschöpfungssysteme (WSS), die Fähigkeiten und Ressourcen mehrerer Organisationen verknüpfen. WSS haben das Potenzial, Lösungen ad-hoc auf konkrete Situationen und Anwendungen zuzuschneiden, um kundenzentrierte und ressourceneffiziente Lösungen zu erbringen [Ni15, Ge18, GK21]. Das Potenzial für verbesserte Angebote kann am Beispiel moderner Autos verdeutlicht werden, die als Plattform für weitere Leistungen bevorstehende Fehler prognostizieren und an eine nahegelegene Werkstatt übermitteln. Die datenbasierte Diagnose ermöglicht bereits vor Eintreffen des Kunden die Personalisierung des angebotenen Leistungsbündels, vom zielgenauen Reparaturprozess, über die Bereithaltung erforderlicher Bauteile bis hin zum Angebot von komplementären Dienstleistungen wie der Bereitstellung eines Mietwagens. Die holistische Perspektive der WSS eröffnet dabei neue Möglichkeiten, sich von der

[1] Universität Hildesheim, Institut für Betriebswirtschaft und Wirtschaftsinformatik, Universitätsplatz 1, 31141 Hildesheim, schoormann@uni-hildesheim.de https://orcid.org/0000-0002-3831-1395
[2] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, friedemann.kammler@dfki.de, https://orcid.org/0000-0002-2418-4882
[3] Leibniz Universität Hannover, Institut für Produktentwicklung und Gerätebau (IPeG), An der Universität 1, 30823 Garbsen, gembarski@ipeg.uni-hannover.de, https://orcid.org/0000-0002-2642-3445
[4] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, simon.hagen@dfki.de

doi:10.18420/inf2022_29

einzelnen innovativen Anwendung zu lösen und unternehmensübergreifende Potenziale zur Verbesserung auch von Nachhaltigkeitszielen umzusetzen.

Der allgemeinen Unterscheidung von Nachhaltigkeit folgend [El97] können Potenziale und Herausforderungen für WSS in die drei Säulen für ökonomische, ökologische und soziale Nachhaltigkeit über den gesamten Lebenszyklus hinweg gegliedert werden.

Vor dem Hintergrund *ökonomischer* Aspekte ergeben sich durch den zunehmenden Bedarf und die Bereitstellung von Daten teils disruptive Veränderungen für den Markt und den Wettbewerb. Kurze Innovationszyklen, die durch den technischen Fortschritt beschleunigt werden, müssen kontinuierlich überwacht werden, um etwa Kompatibilität von Angeboten langfristig sicherzustellen. In diesem Kontext ergeben sich auch neue eigenständige Leistungsangebote, die sich bspw. mit dem automatischen Austausch von Daten zwischen Produkten und Dienstleistungen, aber auch Unternehmen und Kundensegment befassen. Beispiele sind Daten-Marktplätze [Fru20] und Plattform-Geschäftsmodelle [He20, Po22]. Besonders herausfordernd ist hier die Festlegung von Preisen und Werten, die nun dynamisch und kundenspezifisch bestimmt werden können.

Mit dem Fokus auf individuelle Kundenbedürfnisse können Unternehmen und gesamte WSS ihre Aktivitäten so gestalten, dass der Einsatz von Ressourcen (z.B. bei der Produktion) und der Ausstoß von Emissionen (z.B. in der Logistik) minimiert werden. Im Zusammenhang mit *ökologischen* Zielen entstehen durch WSS neue Wege Ressourcen einzusparen und/oder wiederzuwenden. Dazu zählen das Virtualisieren von Produktfunktionen (z.B. Digitale Zwillinge [WD21, St22]), das Ersetzen von Dienstleistungen durch datengetriebene Varianten [KB19, Sc19], das Teilen von Produkten (z.B. Sharing Economy [Sc17, GK21]) oder das Erkunden alternativer Verwendungen von „ausgemusterten“ Leistungen [GL18, Br20].

Neben den ökonomischen und ökologischen Facetten, spielt zunehmend auch die *soziale* Nachhaltigkeitsdimension eine wichtige Rolle [SK20]. Einerseits können neue Infrastrukturen den Zugang zu großen Datenmengen ermöglichen, um auch kleinere Unternehmen und Startups in die Lage zu versetzen digitale Lösungsangebote zu entwickeln und zu vertreiben. Anderseits besteht die Gefahr von sich verhärtenden Marktstrukturen, in denen wenige Organisationen die „Macht“ über Zugänge und Daten haben und somit starke Abhängigkeiten bis hin zu Monopol-artigen Verhältnissen schaffen [HA17]. Fragen hinsichtlich digitaler Verantwortung, transparentem Umgang mit Daten, Schaffung von Vertrauen und Sicherstellung/Monitoring sozial-gerechter Produktions- und Lieferketten gilt es daher zu integrieren.

Aus wissenschaftlicher Perspektive entsteht im Kontext von WSS und Nachhaltigkeit ein interdisziplinäres Spannungsfeld, das unterschiedliche Fragestellungen und Anwendungsfelder verbindet. Hierzu zählt sowohl der Fortschritt datengetriebener Technologien, auch mit Blick auf innovative Anwendungen der Künstlichen Intelligenz [Sc21], als auch grundlegende Infrastrukturmaßnahmen, bspw. der Aufbau von sicheren, digitalen Datenökosystemen wie GAIA-X [Ga22]. Andererseits ist zu klären, wie technische Innovationen messbar und zuverlässig Beiträge zu ökonomisch, ökologisch

und sozial nachhaltigen Angeboten leisten können. Hierfür gilt es, auch das Verhältnis zwischen erzieltem Nutzen und in Kauf genommenen Aufwänden kritisch zu beleuchten und neue, wirksame Geschäfts- und Kooperationsmodelle zu identifizieren.

2 Beiträge: NaWerSys-Workshop 2022

Der NaWerSys-Workshop möchte in dem oben genannten Spannungsfeld Forscherinnen und Forschern die Möglichkeit geben, Forschung und Praxiseinblicke zu präsentieren. Konzeptuelle, empirische und prototypische Beiträge sollen dabei innovative Anwendungen datengetriebener WSS verdeutlichen sowie deren Implementierung in neuen, nachhaltigeren Geschäftsmodellen reflektieren. Ein besonderes Anliegen des Workshops liegt darin, Entwicklungsteams übergreifend zusammenzubringen, erfolgreiche Anwendungen zu diskutieren und gemeinsame Konzepte, Modelle und Methoden zu prägen, die zur systemischen Gestaltung beitragen.

Sechs angenommene Beiträge greifen im zweiten NaWerSys-Workshop, der im Rahmen der INFORMATIK 2022 abgehalten wird, die übergeordnete Motivation auf.

Der Forschungsbeitrag von *Julia Christina Schweihoff*, *Ilka Jussen*, *Maleen Stachon* und *Frederik Möller* trägt zum Verständnis von Geschäftsmodellen in Datenökosystemen bei. Das Autorenteam entwickelt eine Taxonomie, die verschiedene Gestaltungsoptionen für diese Art von Geschäftsmodellen strukturiert.

Jonas Brinker, *Friedemann Kammler* und *Oliver Thomas* nehmen sich der Herausforderung von außerplanmäßigen Instandsetzungen im Maschinen- und Anlagenbau an. Mit dem Service-Tailoring-Mechanismus beantworten die Autoren Fragen in Bezug auf die automatische Anpassung von Serviceangeboten.

Das Autorenteam um *Henrik Kortum*, *Tobias Kohl*, *Dominik Hubertus*, *Oliver Hinz* und *Oliver Thomas* untersucht, wie Machine-Learning-basierte Smart Services in Daten-Ökosystemen eingeführt und betrieben werden können. Basierend auf Smart-Living-Anwendungsfälle wird eine neue Plattform-Architektur präsentiert.

Dem zunehmenden Druck, neue Ideen und Geschäftsmodelle für das Teilen von Daten zu entwickeln, widmen sich *Ilka Jussen*, *Julia Christina Schweihoff*, *Maleen Stachon* und *Frederik Möller*. Der Beitrag präsentiert ein Data Sharing-Toolkit, das mit Hilfe von Canvas-basierten Werkzeugen neue Möglichkeiten für Data Sharing aufzeigt.

Jessica Bollenbach, *Stefan Neubig*, *Andreas Hein*, *Robert Keller* und *Helmut Krcmar* adressieren mit dem Fokus auf die steigenden Emissionen und „Überfüllung“ von Orten aktuelle Herausforderungen im Tourismus. Der Beitrag vergleicht verschiedene Machine-Learning-Algorithmen, um den Andrang und die Überfüllung von lokalen, touristischen Point of Interests vorherzusagen.

Der Beitrag von *Dimitri Petrik*, *Stefan Trieflinger* und *Felix Schönhofen* forciert das Zusammenbringen und Abstimmen unterschiedlicher Interessen, die insbesondere bei

großen Kooperationen in Ökosystemen vorherrschen. Um den Abstimmungsvorgang zwischen den beteiligten Unternehmen zu unterstützen, wird ein Roadmapping-basierter Ansatz entwickelt.

3 Danksagung: NaWerSys-Programmkomitee 2022

Der Workshop wäre nicht möglich gewesen ohne die Unterstützung durch das Organisationsteam der Jahrestagung INFORMATIK 2022. Ein besonderer Dank geht darüber hinaus auch an das NaWerSys-Programmkomitee, das diesen Workshop mit sorgfältig erstellten Gutachten unterstützt hat:

- Dennis Behrens (Data Scientist)
- Katja Bley (Technische Universität Dresden)
- Helge Fischer (FOM Essen)
- Lucas Hüer (HS Osnabrück)
- Marie Kammler (HWG Ludwigshafen)
- Robert Keller (Hochschule Kempten)
- Babett Kühne (Kühne + Nagel)
- Sven Jannaber (LVM Versicherung)
- Cristina Mihale-Wilson (Goethe-Universität Frankfurt)
- Frederik Möller (TU Dortmund Universität)
- Maren Stadtländer (Universität Hildesheim)
- Frederic Stahl (DFKI Oldenburg, Marine Perception)
- Andreas Varwig (Consultant Data Science)
- Sebastian Werning (IU Internationale Hochschule)
- Jannis Vogel (DFKI Osnabrück)

Wir wünschen allen Teilnehmerinnen und Teilnehmern des NaWerSys-Workshops viele neuen Erkenntnisse, spannende Diskussionen und interessante Kontakte.

September 2022
Thorsten Schoormann, Friedemann Kammler, Paul Gembarski und Simon Hagen

Litertaurverzeichnis

[Br20] Brinker, J.; Gembarski, P.C.; Hagen, S.; Thomas, O.: Anwendungspotenziale von Additive Repair und Refurbishment für Service-orientierte Geschäftsmodelle. In Konstruktion für die Additive Fertigung 2019 (pp. 43-54). Springer Vieweg, Berlin, Heidelberg, 2020.

[Di20] Di Vaio, A.; Palladino, R.; Hassan, R.; Escobar, O.: Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. Journal of Business Research 121, S. 283-314, 2020.

[El97] Elkington, J.: The triple bottom line. Environmental management: Readings and cases, 49-66, 1997.

[Fr20] Fruhwirth, M.; Rachinger, M.; Prlja, E.: Discovering business models of data marketplaces. In: Proceedings of the Hawaii International Conference on System Sciences, 2020.

[Ga22] GAIA-X, https://gaia-x.eu/what-is-gaia-x/about-gaia-x/, Access: 04.07.2022

[Ge18] Gembarski, P.C.; Schoormann, T.; Schreiber, D.; Knackstedt, R.; Lachmayer, R.: Effects of mass customization on sustainability: A literature-based analysis. In Customization 4.0 (S. 285-300). Springer, Cham, 2018.

[GK21] Gembarski, P.C.; Kammler, K.: Mass Customizing for Circular and Sharing Economies: A Resource-Based View on Outside of the Box Scenarios. In Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems, S. 1039-1046. Springer, Cham, 2021.

[GL18] Gembarski, P.C.; Lachmayer, R.: Product-Service-Systems – What and why Developers can learn from Mass Customization. Enterprise Modelling and Information Systems Architectures 13, S. 16-1, 2018.

[HA17] Hagiu, A.; Altman, E. J.: Finding the platform in your product. Harvard Business Review 95/4, S. 94-100, 2017.

[Ha19] Hagen, S.; Brinker, J.; Gembarski, P.C.; Lachmayer, R.; Thomas, O.: Integration von Smarten Produkten und Dienstleistungen im IoT-Zeitalter: Ein Graph-basierter Entwicklungsansatz. HMD Praxis der Wirtschaftsinformatik 56, S. 1220–1232, 2019.

[He20] Hein, A.; Schreieck, M.; Riasanow, T.; Setzke D.S.; Wiesche, M.; Böhm, M.; Krcmar, H.: Digital platform ecosystems. Electronic Markets 30, S. 87-98, 2020.

[KB19] Kühne, B.; Böhmann, T.: Data-Driven Business Models-Building the Bridge between Data and Value. In: Proceedings of the European Conference on Information Systems, 2019.

[Lo11] Loos, P.; Nebel, W.; Gómez, J.M.; Hassan, H.; Watson, R.T.; vom Brocke, J.; Seidel, S.; Recker, J.: Green IT: a matter of business and information systems engineering? Business & Information Systems Engineering 3/4, S. 245-252, 2011.

[Lo21] Lobschat, L.; Mueller, B.; Eggers, F.; Brandimarte, L.; Diefenbach, S.; Kroschke, M.; Wirtz, J.: Corporate digital responsibility. Journal of Business Research, S. 122, 2021.

[Ni15] Niemöller, C.; Bärtling, N.; Thomas, O.: Nachhaltigkeit durch Hybride Wertschöpfung-Entwicklung eines Reifegradmodells. In: Proceedings of the INFORMATIK, 2015.

[Po22] Poniatowski, M.; Lüttenberg, H.; Beverungen, D.; Kundisch, D: Three layers of abstraction: a conceptual framework for theorizing digital multi-sided platforms. Information Systems and E-business Management 20/2, S. 257-283, 2022.

[Sc17] Schoormann, T.; Behrens, D.; Knackstedt, R.: Carsharing Geschäftsmodelle–Entwicklung eines bausteinbasierten Modellierungsansatzes. In Smart Service Engineering, S. 303-325. Springer Gabler, Wiesbaden, 2017.

[Sc21] Schoormann, T.; Strobel, G.; Möller, F.; Petrik, D.: Achieving Sustainability with Artificial Intelligence—A Survey of Information Systems Research. In: Proceedings of the International Conference on Information Systems, Austin, USA, 2021.

[Sc19] Schüritz, R.; Farrell, K.; Wixom, B. W.; Satzger, G.: Value co-creation in data-driven services: towards a deeper understanding of the joint sphere. In: Proceedings of the International Conference on Information Systems, 2019.

[SK20] Schoormann, T.; Kutzner, K.: Towards Understanding Social Sustainability: An Information Systems Research-Perspective. In: Proceedings of the International Conference on Information Systems, Hyderabad, India, 2020.

[St22] Strobel, G.; Möller, F.; van der Valk, H.: Healthcare in the era of digital twins: Towards a domain-specific taxonomy. In: Proceedings of the European Conference on Information Systems, Timișoara, Romania, 2022.

[To20] Thomas, O.; Hagen, S.; Frank, U.; Recker, J.; Wessel, L.; Kammler, F.; Zarvic, N.; Timm, I.: Global crises and the role of BISE. Business & Information Systems Engineering, 62/4, S. 385-396, 2020.

[WD21] Wache, H.; Dinter, B.: Digital Twins at the Heart of Smart Service Systems-An Action Design Research Study. In: Proceedings of the European Conference on Information Systems, Virtual, 2021.

Design Options for Data-Driven Business Models in Data-Ecosystems

Julia Christina Schweihoff [1], Ilka Jussen[2], Maleen Stachon[3] and Frederik Möller[4]

Abstract: Data has fundamentally changed the way companies cooperate. Traditional strategies of companies require to be rethought, revisioned, and reimagined based on the new environment of the data economy. Data as a key resource opens up new opportunities that companies need to leverage in data-driven business models (DDBM). Instead of acting in a vacuum, working together and creating value conjointly in an ecosystem based on data is paramount to success. Although DDBMs and ecosystems are not new, per se, there has been a lack of an overarching view of how many DDBMs interact in the context of ecosystems. The paper starts at this point as it develops a taxonomy of these business models in ecosystems. Thus, it contributes to helping researchers and practitioners understand the specifics of data-driven business models working in ecosystems.

Keywords: Data-Driven Business Model, Data Ecosystem, Taxonomy Development, Case Study Research, Design Options.

1 Introduction

Digitalization leads to a fundamental change in the business world, resulting from the increasing availability and importance of data [BO15]. A study by the Cologne Institute for Economic Research [In21] in 2020 finds that companies already attest to the increasing value of data. However, data in its raw state is of little value to companies. Only its meaningful use generates benefits and creates value [FBP20]. This leads to the need to incorporate data into how companies operate, i.e., their business models and how they interact with each other [EGW16]. The faster and easier ways of exchanging information and thus benefiting from the strengths of others have also changed the environment in which companies find themselves. Today, companies often have an extensive network where participants generate value together in so-called *ecosystems* [YCL14, Az20]. The literature discusses various streams of ecosystem research, which usually cannot be concretely distinguished from each other [Gu20a, OL18]. In the context of data-driven business models, data ecosystems are of particular interest because they focus distinctly

[1] Lehrstuhl für Industrielles Informationsmanagement/TU Dortmund, Joseph-von-Fraunhoferstraße 2-4, 44227 Dortmund, julia.schweihoff@tu-dortmund.de, https://orcid.org/0000-0002-2571-8524
[2] Lehrstuhl für Industrielles Informationsmanagement/TU Dortmund, Joseph-von-Fraunhoferstraße 2-4, 44227 Dortmund, ilka.jussen@tu-dortmund.de
[3] Lehrstuhl für Industrielles Informationsmanagement/TU Dortmund, Joseph-von-Fraunhoferstraße 2-4, 44227 Dortmund, maleen.stachon@tu-dortmund.de
[4] Lehrstuhl für Industrielles Informationsmanagement/TU Dortmund, Joseph-von-Fraunhoferstraße 2-4, 44227 Dortmund; Fraunhofer ISST, Emil-Figge-Str. 91, 44227 Dortmund, frederik.möller@tu-dortmund.de

doi:10.18420/inf2022_30

on data as the focal object (e.g., [OL18]). A study by the European Commission shows that the integration and meaningful use of data in ecosystems is essential to break down existing data silos and prevent the formation of future silos in the interests of sustainability and ecosystem self-sustainability [Ma21]. The use of data can also contribute to the achievement of sustainability goals. For example, research by [Sc21] shows that in many cases, access to data sets lays the foundation for achieving sustainability goals.

The paper aims to assist companies and researchers in grasping the complexity of data-driven business models working in data ecosystems. For this purpose, we propose a taxonomy of these data-driven business models. A taxonomy is a way of classifying, sorting, and structuring knowledge to understand a specific subject area [GV95, NVM13, SSK19]. Taxonomies originate in biology and have experienced high popularity in Information Systems (IS) research [NVM13], in particular, through their importance to business model research (e.g., [La15, KH12]). Business model taxonomies are a prevailing artifact that supports managers and researchers in understanding, analyzing, and designing new business models (e.g., [La15]). [Mö21] provide an overview of existing business model taxonomies based on the domains they cover (e.g., logistics [Mö20] or car-sharing [Re16]) and the technologies they use (e.g., digital platforms [FRP20b]). Subsequently, while business model taxonomies exist, there is yet none that sheds light on an increasingly relevant field, i.e., data-driven business models in ecosystems. This research aims to help companies understand the complexity of synergies in an ecosystem to achieve an optimal alignment of their business models. Based on the above, we formulate the following research question:

Research question: *What are the design options for data-driven business models in a data ecosystem environment?*

We pursue this research question to understand how a DDBM should operate in the environment of a data ecosystem. Based on understanding a data-driven business model in an ecosystem environment, the next step is to analyze the existing literature to derive design options. This paper is structured as follows. First, we present the background of our work, i.e., data-driven business models and data ecosystems. Next, we present the research approach we took to designing the taxonomy. The fourth section of the paper presents the final taxonomy and its specific manifestations. Section 5 discusses the results and Section 6 concludes with an overview of the contributions, limitations, and an outlook for future work.

2 Theoretical Background

2.1 Data-driven business models

Although the term “business model” has been around since the 1950s, interest did not surge until the late 1990s [Os04]. [AEA08, S. 1] highlights that the concept of business

models has evolved into "one of the most important domains in the field of Information Systems (IS), thanks to recent rapid advances in Information and Communication Technologies (ICTs)" with "(…) as many definitions as there are business models" [Te18, S. 41]. Thus, the understanding of what makes a business model is not consistent. However, for this paper, we follow the general definition of [Te10, S. 173], who defines a business model as "the logic (…) how a business creates and delivers value to customers" and, further, as "the architecture of revenues, costs, and profits associated with the business enterprise delivering that value". Sometimes, a subrange of a business model is understood under the term "business model", which does not represent a holistic business model [LC00, Os04]. According to [Os04], it is not sufficient to look at individual parts of a business model. Instead, it is crucial for success to see the complete picture. Likewise, the potential for success of a business model must not be taken for granted. Ultimately, success depends on the implementation and execution of those responsible [Os04].

Motivated by the increasing amount of data in business, new business model streams have emerged investigating data-driven business models. Given the nature and timeliness of the topic, DDBMs are still a young and emerging field in IS and business model research [KB19, Gu20b, Ha16, Ha20]. The literature does not yet provide a clear definition of a data-driven business model [EGW16, Mö20]. For example, the terms "data-driven" and "data-based" can be found in the literature [SS16]. In this paper, we see these terms as equivalent and, in the following, will only use "data-driven" business models. Contrary to similar terms, DDBM definitions have a shared focus: the use of data as the core resource in the business model (e.g., [Mö20, Ha16]).

2.2 Data ecosystems

[OL18] define a data ecosystem as a multitude of networks consisting of autonomous actors that directly or indirectly consume, offer, or provide linked resources, such as software or infrastructure, to data. Actors take one or more roles within the networks and are linked through relationships with other actors to maintain collaboration but also competition within a data ecosystem [OBF19]. According to [OL18], data ecosystems consist of the following four elements: *Actors*, *Roles*, *Relationships*, and *Resources*.

In an ecosystem, an actor can be a company or an individual who takes on one or more roles. It must be able to fulfill requirements placed on its role by other participants in the ecosystem [OBF19]. Basically, actors have different expectations and other motivational drivers that are reconciled in terms of the data ecosystem [OL18]. [OL18] define a role as a function that is linked to a variety of tasks and activities. At least the data consumer and the data producer typically exist in a data ecosystem. In addition, there may be other roles, such as the intermediary, which may overlap in their responsibilities. The tasks of a producer include data selection and the preparation, maintenance, and publication of data sets [OBF19]. On the other hand, consumers gain access to the datasets, analyze them, and provide feedback if desired. For the role of the intermediary, tasks such as developing and providing solutions and maintaining them are envisaged [OL18]. The relationship aspect

covers the interaction between different actors based on common interests or depending on economic or technological reasons. For example, data or other resources are exchanged via transactions, and, ideally, new business models can be derived from such relationships [OL18].

According to [OL18], resources are products or capabilities produced, provided, or consumed by actors. In a data ecosystem, resources can be data sets on the one hand and data-based software or infrastructure on the other. The resources are defined by fixed standards or licenses and can be transferred between actors both individually and in combination [OL18]. Overall, benefits can be derived from establishing or developing a data ecosystem beyond political, social, and economic factors. These include, for example, an improvement in the quality of data and services or also the increased exchange between the actors [OBF19]. In general, activities such as the joint generation, processing, and use of data for analyses form the core of a data ecosystem and thus represent a clear added value for the participating actors [Az20]. Reasons for deciding against this development include aspects such as a lack of knowledge around ecosystems, the high level of complexity in dealing with data, and the lack of actors for an ecosystem [OBF19].

3 Research Design

In our research, we propose a taxonomy. The foundation of the taxonomy is a systematic literature review using the terms „data-driven business models“ and „data ecosystem“. The keywords were developed according to the three-stage procedure by [Sc18]. To develop the taxonomy, we use the method of [NVM13], which is the *de facto* standard in developing business model taxonomies (e.g., see [Mö21]). In a first step, we searched three databases, i.e., Google Scholar, Scopus, and AISeL, for the terms “data-driven business model” and “data ecosystem”. We filtered the initial corpus of literature based on [WW02] and generated a sub-corpus to include in the taxonomy design (see Figure 1).

According to [NVM13], the usability of a taxonomy is confirmed by its applicability. Therefore, practical examples from the Gaia-X environment were used to test whether the taxonomy is applicable for an iteration of the taxonomy. Gaia-X is a European initiative that aims to enable secure data sharing. Gaia-X cases were chosen for the practical examples because they are dedicatedly data ecosystems [GA21]. Furthermore, we used the taxonomy evaluation framework by [SSK19] to evaluate the taxonomy through various workshops reviewing the taxonomy. In total, we required five iterations until all ending conditions were met (see Fig. 1). The figure’s representation is inspired by [DGR21].

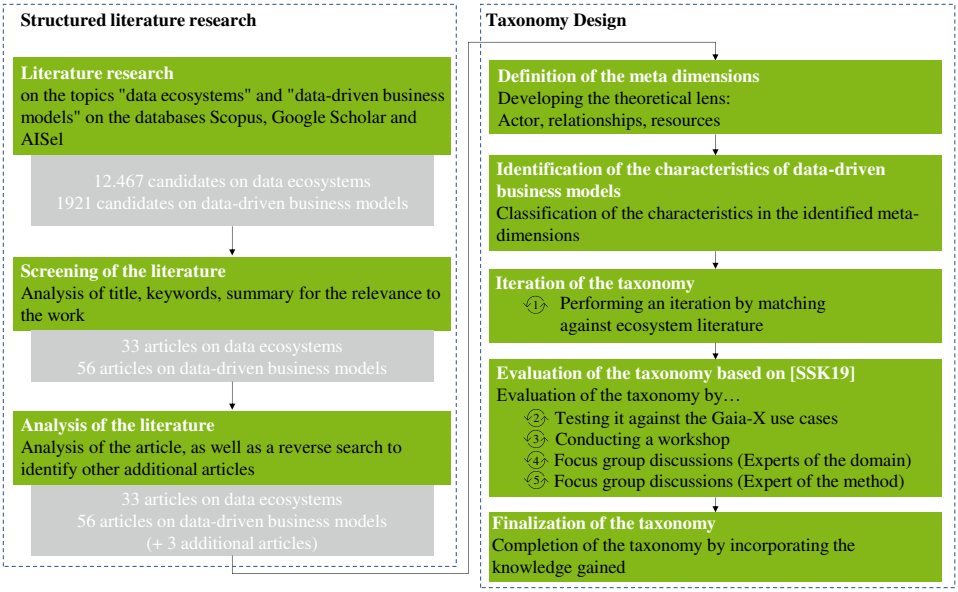


Fig. 1. Taxonomy building approach

4 Design Options of Data-Driven Business Models in Ecosystems

Since our goal is to assist others in finding design options for these business models, the *meta-characteristic* of the taxonomy is as follows: *"Provide design options for a data-driven business model in the environment of a data ecosystem."*

Thus, one goal is to first analyze the corresponding characteristics from the literature. To achieve this goal, the core characteristics of a data ecosystem were first identified. Based on the literature, the aspects of actors, roles, relationships, and resources represent the core components of a data ecosystem that we use as meta-dimensions (e.g., [Mö21]).

- *Actor:* What role does the company play in the ecosystem?
- *Relationship:* How does the interaction with partners and customers take place?
- *Resources:* What resources are used to generate value?
- *Value:* How is value created for the company itself and its customers?

The final taxonomy is shown in Fig. 2. To make the taxonomy more tangible, two examples of possible expressions are also given in Figure 2. These were filled in during the evaluation using the Gaia-X use cases. The basis for this is the Agri-Gaia [BM21] and the Collaborative Condition Maintenance use cases [Bu21].

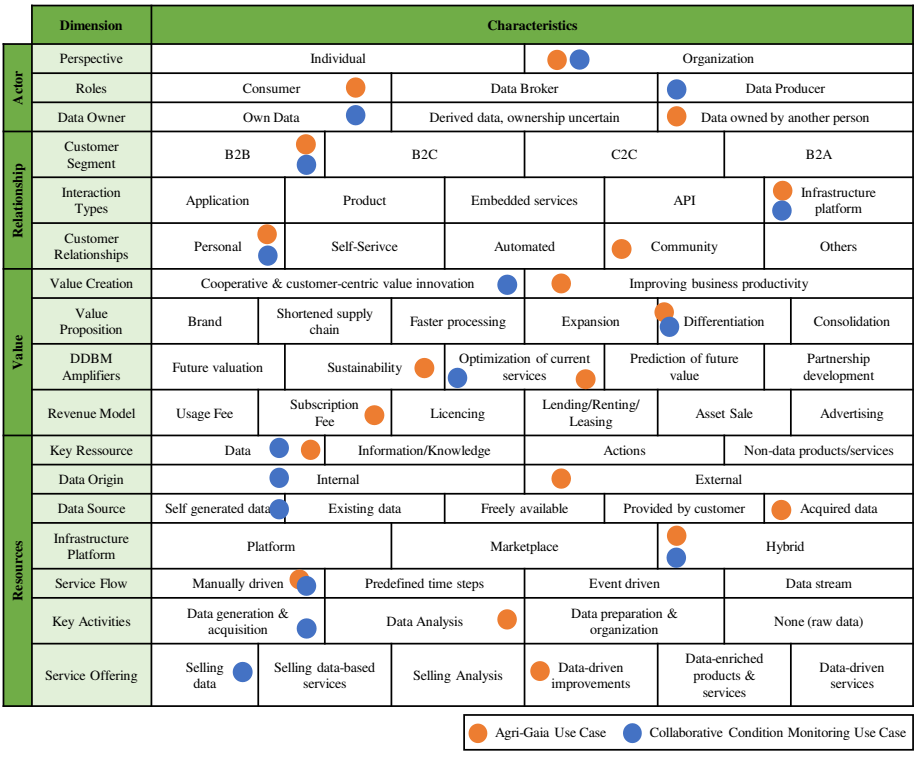


Fig. 2. Taxonomy of data-driven business models in ecosystems

4.1 Meta-Dimension: Actor

A pivotal decision for data ecosystem participants is deciding on how to participate. They must first determine what **perspective** they want to take in the data ecosystem. They can choose between an *individual* perspective or an *organization* [Ol18]. Once actors are clear about their capabilities and motivations, the last thing they must decide is which **role** they want to take. Choosing a role is not exclusive, as actors can also take on several roles simultaneously [OL18, Ge21]. For example, if the company only wants to receive data to leverage them in its business model, it should choose the role of the *data user* [BTS14, OL18, Ol18]. If an actor produces data and wants to make it available to other companies, it will find itself in the role of *data producer* [Ol18, OL18, BTS14]. The third identified role represents that of the *data broker* [BTS14]. This person or company is responsible for

distributing the data between the different actors and assigning user permissions. In this way, this role supports the other actors using the data. The data broker thus takes his place between the data user and the data producer, thus linking these two roles.

4.2 Meta-Dimension: Relationship

The consideration and subsequent selection of a targeted **customer segment** represent another important point connected to business model elaboration. By selecting a customer segment, the company can tailor its value proposition precisely to its target customers and generate maximum value [DGR21, Os04]. In many consistent cases, the two customer segments identified by [Ha14] are classically mentioned at this point in the literature: *B2B* (business-to-business) and *B2C* (business-to-customer). The traditional segmentation of customer groups is supplemented by the segments *C2C* (customer-to-customer) [Br15] and *B2A* (business-to-administrative) [DGR21]. Once an organization has decided on a customer segment, it can plan the **type of interaction** with the corresponding customer segment based on this. According to [DGR21], the interaction type shows how the customer receives the actual product value. Based on [DGR21], there are four interaction types: *Application, Product, Embedded Services* [RBE18], and *Application Programming Interface (API)* [Mö20].

In addition to the choice of interaction type, the **customer relationship**'s type and nature also determine the business model. The type of customer relationship refers to how the customer receives the product or data. This does not mean whether the customer receives a product or an application (cf. interaction type), but whether they interact with their business partner, for example, through personal contact or in an automated manner. This aspect is crucial for the business model's long-term success in cultivating and permanently maintaining customer and business partner relationships [DGR21]. Further, [DGR21] states that there are four specific ways of customer interaction. Based on the findings of [DGR21], the listed interaction types are also adopted as possible interaction types in this research. These include *personal interaction*, *self-service*, *automated interactions*, *communities*, or *other* variants that are not clearly defined.

4.3 Meta-Dimension: Value

To generate revenue with a product, service, or similar, a **revenue model** must first be selected. This dimension represents a highly crucial point in a data-driven business model, as the evaluation of the relevant literature shows. 46 out of 58 data sets deal with or at least mention the topic of the **revenue model**. Specifically, the revenue models *Usage Fee*, *Subscription Fee*, *Licencing*, *Lending/Renting/Leasing*, *Asset Sale,* and *Advertising* are dealt with [BN18, Br15, DGR21, Ha14, KB18, SSD17, Za19]. Furthermore, when choosing a business model, it must be defined how exactly value is created with the product. This question is also very central to the profit of the business model. In their research, [ZCG16] examined various patterns of **value generation**. They identified four

possible variants that could be considered for this purpose: cooperative value innovation, customer-centric value innovation, cooperative productivity improvement, and company-centric production improvement. Finally, further analyses have shown that there is a combination of the different variants in most cases. Thus, it can be summarized that mostly either *cooperative and customer-centric value innovation or enterprise* and *cooperative productivity improvement* are present [ZCG16]. The **value proposition** is closely related to the way value is created. The selection of characteristics mentioned here is based on a study by [Br15]. One evaluation point concerned the intended goal of the companies to be achieved using data. Consequently, the following characteristics emerge: *Brand, shortened supply chain, faster processing, expansion, differentiation,* and *consolidation* [Br15]. **Business model amplifiers, or DDBM amplifiers**, provide information on how a company plans to establish its business model in the long term. In their research article, [Is20] identified several key factors critical to the success of a company and its data-driven value creation. Overarchingly, these are summarized under the term's optimization of *current services*, *prediction of future value*, and *development of partnerships*. They are additionally supplemented by the aspect of *sustainability* [BM20, BTS14, ESK17, Is20, LD20, MB19, FRP20a]. The focus of the sustainability characteristic is the analysis of how a company can achieve longer life cycles but also recycle different components or products [ESK17]. Business model enhancers are rounded off by their future evaluation in the form of a shared *vision* or a unique selling proposition [ESK17].

4.4 Meta-Dimension: Resources

Key resources are another core component and represent what the company delivers to the customer as a core product [DGR21]. A data-driven business model includes *data*, *information and knowledge*, *actions,* and *non-data products or services* [Ha14]. This is a continuous further development of the data in each step. [Ha14] identified seven **data sources** through their research. These data can have an *internal or external origin* [DGR21, Ha16]. If the **origin** is internal, the data is owned by the company. On the one hand, it can be data that is already present in the existing IT systems (*existing data*) or that is *generated through use* but has not yet been used (*self-generated data*). On the other hand, external data is data that has not been generated within the company. This can be acquired data, data *provided by the customer*, or *freely available* data [Ha14]. The process of how the product reaches the consumer describes the **service flow**. The consumer can receive the offered product *manually*, in *predefined time steps*, triggered by certain *events*, or in a *stream* [DGR21]. The various manifestations differ primarily in the activity of the individual participants.

The service flow is closely related to the service offering. In a data-driven business model, the **service offering** represents the sale of data, sale of *data-based services*, *sale of analytics*, *data-driven improvements*, *data-enriched products and services*, and *data-enabled services* [Br19, Pa20]. **Key activities** form a crucial part of a business model. They represent the activities that a company performs to be able to produce the planned service offering and then deliver it to its customers [Ha14]. Key activities are what give

data its true value [Br15]. Since data does not represent a product in the classic sense in the physical world and the classic value chain sees information only as a supporting element and not as a source of value, the view must be adapted at this point [Ha14]. The key activities identified in the taxonomy include *data generation and collection*, *data analysis*, *data preparation and organization*, and *raw data* that has not undergone any processing [Br15, DGR21, Ha14].

5 Contributions, Limitations, and Outlook

The importance of data is increasing due to digitalization. The resulting influence leads to a change in structures from traditional business models to business models that use data as their key resource [BO15]. Companies that use data as their key resource have a data-driven business model. They stand out by not only owning the data but using it to generate value for themselves and their customers [EGW16, Gu20b, Ha14, KB18, SSD17, ZAG17]. The increasing focus on data is not the only side effect of rising digitization. One reaction is a change in the way companies work together. Instead of acting on their own as before, so-called ecosystems are forming. These ecosystems require new thinking within companies [FJR19], which we analyzed in this paper codified as a taxonomy.

As a result, new opportunities for interaction and marketing of new products emerge, ultimately sustained by participants making the best possible contribution to the ecosystem based on their intrinsic motivation [OL18, Ol18]. Now, to derive design options for data-driven business models in ecosystems, an overview of the basic components of a data ecosystem is first needed. Looking at the research by [Ol18] and [OL18], one recognizes four main aspects that make up a data ecosystem: Actors, Roles, Relationships, and Resources. These four core elements form the basis for a data-driven business model in an ecosystem. In addition, other characteristics are relevant that primarily relate to the business model in general. These include the value proposition, the required interfaces, the service platform, the organizational model, and the revenue model [EP13]. Starting from the beginning of a company's participation or realignment in an ecosystem, the first step is to position itself in the ecosystem.

For this, the company must be aware of its motivation as well as its capabilities, which our taxonomy assists. Motivation and capabilities form the basis on which the company selects its role in the ecosystem. Then, value generation, relationships, and resources need to be defined. The relationships define the interaction paths and the associated organizational aspects, such as selecting customer segments or even the manner of customer relations. The resources and the value generation explain how the company can generate monetarily or also other individual values for itself and/or its customers through the resources at its disposal. In addition, aspects such as the key activities and the range of services are defined, along with the key resources, forming a core component of the business model.

While evaluating the developed taxonomy based on a literature review and the exemplary application of various testers, it became clear that it is suitable for representing data-driven

business models in ecosystems. Obstacles that became apparent during the evaluation iterations were modified step by step to enable the taxonomy to be applied intuitively. What emerged from this is knowledge about the versatility of DDBMs. From different data sources to various value propositions to diverse service offerings, DDBMs can differ significantly.

This research is subject to various limitations. First, further research projects with the same research objective may lead to different results. The reason for this is that the evaluation of the taxonomy is subject to the subjective point of view of the researcher. The same applies to the evaluations carried out. Furthermore, the literary basis is based on reviewing the databases Scopus, AISeL, and Google Scholar. Using other databases or conducting the literature search at an earlier or later date may identify other research articles, which would produce additional results. Another limiting aspect is the search terms used. For example, more specific or broader search terms will yield different results in the literature search, which would also have implications for the results. Likewise, further research can be action guidelines for developing or adapting data-driven business models to an ecosystem. Conceivable at this point would be using a visual inquiry tool, such as a canvas, which offers companies support in visualizing requirements for these types of business models. In this area, there are already basic variants, such as the Business Model Canvas by [OP10] but not yet a version that relates explicitly to data-driven business models in ecosystems. The taxonomy is also a starting point for others to enrich our findings. For example, by conducting interviews, it is possible to derive recommendations for action based on the explanations provided by the companies interviewed and thus support newly emerging companies.

The taxonomy is also a starting point for others to enrich our findings. For example, by conducting interviews, it is possible to derive recommendations for action based on the explanations provided by the companies interviewed and thus support newly emerging companies. The topic of data-driven business models and the connection to the structures of an ecosystem still holds a lot of potential. This is shown by the small number of research articles addressing this. Even if this work offers a further contribution, especially deeper topics such as data sharing and role constructs within an ecosystem are promising starting points for the future. Furthermore, it is essential to offer companies assistance on how they can adapt their business models even better to the resource of data. Many companies are critical of the sharing of data due to trust issues or a lack of transparency of the options available to date [FPT19, GO20]. It is a task of science to create incentives for data exchange and to show companies the possibilities of how this data exchange can be integrated into their business model and how they can interact in an ecosystem.

6 Acknowledgements

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the Incentives and Economics of Data Sharing Funding Action (IEDS0001). The author is responsible for the content of this publication.

Bibliography

[AEA08] Al-Debei, M. M.; El-Haddadeh, R.; Avison, D.: Defining the Business Model in the New World of Digital Business. In: Proceedings of the Americas Conference on Information Systems, Toronto: Canada, 2008.

[Az20] Azkan, C. et al.: SERVICE DOMINANT LOGIC PERSPECTIVE ON DATA ECOSYSTEMS - A CASE STUDY BASED MORPHOLOGY. In: Proceedings of the 28th European Conference on Information Systems, 2020.

[BM20] Breidbach, C. F.; Maglio, P.: Accountable algorithms? The ethical implications of data-driven business models. Journal of Service Management 2/31, pp. 163–185, 2020.

[BM21] BMWi: Agri-Gaia. https://www.bmwi.de/Redaktion/DE/Artikel/Digitale-Welt/GAIA-X-Use-Cases/agri-gaia.html, Last Access: 02.09.2021.

[BN18] Böhmecke-Schwafert, M.; Niebel, C.: THE GENERAL DATA PROTECTION'S (GDPR) IMPACT ON DATA-DRIVEN BUSINESS MODELS:THE CASE OF THE RIGHTTODATA PORTABILITY AND FACEBOOK. ITU Journal: ICT Discoveries, Special Issue No. 2, 2018.

[BO15] Barenfanger, R.; Otto, B.: Proposing a Capability Perspective on Digital Business Models. In: IEEE 17th Conference on Business Informatics. IEEE, pp. 17–25, 2015.

[Br15] Brownlow, J. et al.: Data and Analytics - Data-Driven Business Models: A Blueprint for Innovation, 2015.

[Br19] Breitfuss, G. et al.: The Data-Driven Business Value Matrix - A Classification Scheme for Data-Driven Business Models. In (Pucihar, A.; Kljajić Borštnar, M.; Vidmar, D. eds.): Humanizing technology for a sustainable society. Maribor, pp. 804–819, 2019.

[BTS14] Bulger, M.; Taylor, G.; Schroeder, R.: Data-driven business models: challenges and opportuities of big data. Oxford Internet Institute. Research Councils UK: NEMODE, New Economic Models in the Digital Economy, 2014.

[Bu21] Bundesministerium für Wirtschaft und Energie: Collaborative Condition Monitoring. https://www.bmwi.de/Redaktion/DE/Artikel/Digitale-Welt/GAIA-X-Use-Cases/collaborative-condition-monitoring.html, Last Access: 13.09.2021.

[DGR21] Dehnert, M.; Gleiss, A.; Reiss, F.: What makes a data-driven business model? A consolidated taxonomy. In: Proceedings of the European Conference on Information Systems, virtual, 2021.

[EGW16] Engelbrecht, A.; Gerlach, J.; Widjaja, T.: Understanding the anatomy of data-driven business models. Towards an empirical taxonomy. In: Proceedings 24th European Conference on Information Systems, Istanbul: Turkey, 2016.

[EP13] El Sawy, O. A.; Pereira, F.: Business Modelling in the Dynamic Digital Space an Ecosystem Approach. Springer, Berlin, 2013.

[ESK17] Exner, K.; Stark, R.; Kim, J. Y.: Data-driven business model. A methodology to develop smart services. In: 23rd International Conference on Engineering, Technology and Innovation: Engineering, Technology and Innovation, pp. 146–154, 2017.

[FBP20] Fruhwirth, M.; Breitfuss, G.; Pammer-Schindler, V.: The Data Product Canvas a visual collaborative tool for designing data-driven business models. In: Proceedings of the BLED, pp. 516–528, 2020.

[FJR19] Fuller, J.; Jacobides, M. G.; Reeves, M.: The Myths and Realities of Business Ecosystems. MIT Sloan Management Review 3/60, pp. 1–9, 2019.

[FPT19] Fruhwirth, M.; Pammer-Schindler, V.; Thalmann, S.: To Sell or Not to Sell: Knowledge Risks in Data-Driven Business Models. In: Proceedings of the 2019 Pre-ICIS SIGDSA Symposium 11, pp. 1–9, 2019.

[FRP20a] Fruhwirth, M.; Ropposch, C.; Pammer-Schindler, V.: Supporting Data-Driven Business Model Innovations. Journal of Business Models, Vol. 8 No. 1 (2020): Online First of Special Issue: Fostering Multi-and Interdisciplinary Business Model Research / Journal of Business Models, Vol. 8 No. 1 (2020): Online First of Special Issue: Fostering Multi- and Interdisciplinary Business Model Research, 2020.

[FRP20b] Fruhwirth, M.; Rachinger, M.; Prlja, E.: Discovering Business Models of Data Marketplaces. In: Proceedings of the 53rd Hawaii International Conference on System Sciences, pp. 5738–5747, 2020.

[GA21] GAIA-X: What is Gaia-X. www.gaia-x.eu/what-is-gaia-x, Last Access: 01.09.2021.

[Ge21] Gelhaar, J. et al.: Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. In: Proceedings of the Pacific Asia Conference on Information Systems, pp. 1–14, 2021.

[GO20] Gelhaar, J.; Otto, B.: Challenges in the Emergence of Data Ecosystems. In: Proceedings of the 23rd Pacific Asia Conference on Information Systems, 2020.

[Gu20a] Guggenberger, T. M. et al.: Ecosystem Types in Information Systems. In: Proceedings of the European Conference on Information Systems, 1-21, 2020.

[Gu20b] Guggenberger, T. et al.: Towards a Unifying Understanding of Digital Business Models. 23rd Pacific Asia Conference on Information Systems, pp. 1–14, 2020.

[GV95] Glass, R. L.; Vessey, I.: Contemporary Application-Domain Taxonomies. IEEE Software 4/12, pp. 63–76, 1995.

[Ha14] Hartmann, P. M. et al.: Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms, 2014.

[Ha16] Hartmann, P. M. et al.: Capturing Value from Big Data – A Taxonomy of Data-Driven Business Models Used by Start-Up Firms. International Journal of Operations & Production Management 10/36, pp. 1382–1406, 2016.

[Ha20] Hahn, C. et al.: Exploring AI-Driven Business Models: Conceptualization and Expectations in the Machinery Industry. In: IEEE International Conference on Industrial Engineering and Engineering Management, pp. 567–570, 2020.

[In21] Institut der deutschen Wirtschaft Köln: Datenwirtschaft in Deutschland. Wo stehen die Unternehmen in der Datennutzung und was sind ihre größten Hemmnisse? https://bdi.eu/publikation/news/datenwirtschaft-in-deutschland/, Last Access: 06.01.2022.

[Is20] Isabelle, D. et al.: The role of analytics in data-driven business models of multi-sided platforms: An exploration in the food industry. Technology Innovation Management Review Volume 10, Issue 7, pp. 4–15, 2020.

[KB18] Kühne, B.; Böhmann, T.: Requirements for Representing Data-Driven Business Models - Towards Extending the Business Model Canvas. In: Proceedings 24th Americas Conference on Information Systems, New Orleans: USA, 2018.

[KB19] Kühne, B.; Böhmann, T.: DATA-DRIVEN BUSINESS MODELS – BUILDING THE BRIDGE BETWEEN DATA AND VALUE. In: Proceedings of the European Conference on Information Systems, pp. 1–16, 2019.

[KH12] Kamprath, M.; Halecker, B.: A Systematic Approach for Business Model Taxonomy-How to operationalize and compare large Quantities of Business Models?: ISPIM Innovation Symposium, 2012.

[La15] Lambert, S.: The Importance of Classification to Business Model Research. Journal of Business Models 1/3, pp. 49–61, 2015.

[LC00] Linder, J. C.; Cantrell, S.: Changing Business Models: Surveying the Landscape. Working Paper. Accenture Institute for Strategic Change, 2000.

[LD20] Lange, H. E.; Drews, P.: From ideation to realization: Essential steps and activities for realizing data-driven business models. In: Proceedings - 2020 IEEE 22nd Conference on Business Informatics, CBI 2020, pp. 1–10, 2020.

[Ma21] Martin, S. et al.: Establishment of sustainable data ecosystems. Recommendations for the evolution of spatial data infrastructures. Publications Office of the European Union, Luxembourg, 2021.

[MB19] Müller, J.; Buliga, O.: Archetypes for data-driven business models for manufacturing companies in Industry 4.0. In: International Conference on Information Systems, Special Interest Group on Big Data Proceedings 2, pp. 1–8, 2019.

[Mö20] Möller, F. et al.: Data-driven Business Models in Logistics: A Taxonomy of Optimization and Visibility Services. In: Proceedings of the 53rd Hawaii International Conference on System Sciences, Hawaii, pp. 5379–5388, 2020.

[Mö21] Möller, F. et al.: Designing business model taxonomies – synthesis and guidance from information systems research. Electronic Markets, 2021.

[NVM13] Nickerson, R. C.; Varshney, U.; Muntermann, J.: A Method for Taxonomy Development and its Application in Information Systems. European Journal of Information Systems 3/22, pp. 336–359, 2013.

[OBF19] Oliveira, M. I. S.; Barros Lima, G. d. F.; Farias Lóscio, B.: Investigations into Data Ecosystems: a systematic mapping study. Knowledge and Information Systems 2/61, pp. 589–630, 2019.

[Ol18] Oliveira, M. I. S. et al.: Towards a meta-model for data ecosystems. dg.o '18: dg.o 2018. In: Proceedings of the 19th Annual International Conference on Digital Government Research, Delft, Netherlands. ACM, NewYork, NY, USA, 10 pages., pp. 1–10, 2018.

[OL18] Oliveira, M. I. S.; Lóscio, B. F.: What is a data ecosystem? dg.o '18: dg.o 2018. In: Proceedings of the 19th Annual International Conference on Digital Government Research, Delft, Netherlands. ACM, New York, NY, USA, 9 pages, pp. 1–9, 2018.

[OP10] Osterwalder, A.; Pigneur, Y.: Business Model Generation: Ein Handbuch für Visionäre, Spielveränderer und Herausforderer. Campus Verlag, Frankfurt, New York, 2010.

[Os04] Osterwalder, A.: THE BUSINESS MODEL ONTOLOGY. A PROPOSITION IN A DESIGN SCIENCE APPROACH, Lausanne, France, 2004.

[Pa20] Parvinen, P. et al.: Advancing Data Monetization and the Creation of Data-based Business Models. Communications of the Association for Information Systems, 2020.

[RBE18] Risk, A.; Bergvall-Kåreborn, B.; Elragal, A.: Towards a taxonomy for data-driven digital services. In: Proceedings of the 51st Hawaii international conference on system sciences, pp. 1076–1085, 2018.

[Re16] Remane, G. et al.: A Taxonomy of Carsharing Business Models. In: Proceedings of the 37th International Conference on Information Systems, Dublin, Ireland, 2016.

[Sc18] Schoormann, T. et al.: Sorry, Too Much Information Design Principles for Supporting Rigorous Search Strategies in Literature Reviews. In: IEEE 20th Conference on Business Informatics (CBI), Vienna, Austria, 2018.

[Sc21] Schoormann, T. et al.: Achieving Sustainability with Artificial Intelligence - A Survey of Information Systems Research. In: Proceedings of the 42nd International Conference on Information Systems, pp. 1–17, 2021.

[SS16] Schuritz, R.; Satzger, G.: Patterns of Data-Infused Business Model Innovation. In: IEEE 18th Conference on Business Informatics (CBI). IEEE, pp. 133–142, 2016.

[SSD17] Schüritz, R.; Seebacher, S.; Dorner, R.: Capturing Value from Data: Revenue Models for Data-Driven Services. In: Proceedings of the 50th Hawaii International Conference on System Sciences, pp. 5348–5357, 2017.

[SSK19] Szopinski, D.; Schoormann, T.; Kundisch, D.: Because your Taxonomy of Worth It: Towards a Framewotk for Taxonomy Evaluation. In: Proceedings of the 27th European Conference on Information Systems, Stockholm & Uppsala, Sweden, 2019.

[Te10] Teece, D. J.: Business Models, Business Strategy and Innovation. Long Range Planning 2-3/43, pp. 172–194, 2010.

[Te18] Teece, D. J.: Business Models and Dynamic Capabilities. Long Range Planning 1/51, pp. 40–49, 2018.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 2/26, pp. xiii–xxiii, 2002.

[YCL14] Yoo, S.; Choi, K.; Lee, M.: Business Ecosystem and Ecosystem of Big Data. In (Chen, Y. et al. eds.): Web-Age Information Management. Springer International Publishing, Cham, pp. 337–348, 2014.

[Za19] Zaki, M.: Digital transformation: harnessing digital technologies for the next generation of services. Journal of Services Marketing 4/33, pp. 429–435, 2019.

[ZAG17] Zolnowski, A.; Anke, J.; Gudat, J.: Towards a Cost-Benefit-Analysis of Data-Driven Business Models. In: 13th International Conference on Wirtschaftsinformatik, St. Gallen, Switzerland, pp. 181–195, 2017.

[ZCG16] Zolnowski, A.; Christiansen, T.; Gudat, J.: Business Model Transformation Patterns of Data-Driven Innovations. In: Proceedings of the 24th European Conference on Information Systems, Istanbul, Turkey, 2016.

Service Tailoring: Ein Konfigurationsmechanismus für individualisierte After-Sales-Services im Maschinen- und Anlagenbau

Jonas Brinker [1], Friedemann Kammler[2] und Oliver Thomas[1,2]

Abstract: Außerplanmäßige Instandsetzungen sind eine Herausforderung im Maschinen- und Anlagenbau. Dies betrifft sowohl Nutzer der Maschinen, die aufgrund der Ausfallzeit Abweichungen von ihren Produktionsplänen hinnehmen, aber auch Hersteller, die Serviceressourcen kurzfristig und dynamisch zuordnen müssen. Der Einsatz technischer Innovationen eröffnet dabei neue Serviceoptionen, die beispielsweise vorteilhaft hinsichtlich Zeitaufwands, Kosten oder Nachhaltigkeit sind. Gleichzeitig variieren die kundenindividuellen Anforderungen an die Serviceerbringung, sowohl hinsichtlich der benötigten Ressourcen als auch hinsichtlich individueller Präferenzen. Mit dem „Service Tailoring" (ST) stellen wir einen Mechanismus vor, wie Serviceoptionen kundenspezifisch konfiguriert werden können. Aus elf Experteninterviews erheben wir zunächst Prozessvarianten in der Instandhaltung und bilden einen generalisierten Lösungsraum. Indem wir Mechanismen der Mass-Customization (MC) als Constraint-Optimization-Problem adaptieren, entwickeln wir einen zweistufigen methodischen Vorschlag und demonstrieren diesen an einem Beispiel aus der Landtechnik.

Keywords: Tailored Service, Aftermarket, Data-Driven Service Systems, Service Konfiguration

1 Einleitung

Globalisierte Märkte treiben die kontinuierliche Weiterentwicklung von Produktionsketten zu Wertschöpfungssystemen, die Geschwindigkeit, Qualität und Ressourceneinsatz optimieren [JCG18]. Allerdings wird dabei eine zunehmende Fragilität in Kauf genommen, die zum Beispiel darin beobachtet werden kann, dass gerade in komplexen, eng getakteten Produktionsumgebungen ungeplante Ausfälle zu hohen Kosten führen [Th21]. Dieses Risiko begründet eine Vielzahl technisch und organisatorisch getriebener Service-Innovationen, die die kontinuierliche und tiefgreifende digitale Transformation des Aftermarkets (hier im Sinne der Bereitstellung von Ersatzteilen, Instandhaltungsleistungen sowie Retrofits und Rückführung) vorantreiben. Um Ausfälle von bekannten Verschleißteilen zu mindern, werden beispielsweise präventive und prädiktive Servicestrategien entwi-

[1] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, jonas.brinker@dfki.de, https://orcid.org/0000-0002-5897-1841, friedemann.kammler@dfki.de, https://orcid.org/0000-0002-2418-4882, oliver.thomas@dfki.de
[2] Universität Osnabrück, Fachgebiet Informationsmanagement und Wirtschaftsinformatik, Parkstraße 40, 49080 Osnabrück, oliver.thomas@uni-osnabrueck.de

doi:10.18420/inf2022_31

ckelt. Präventive Strategien senken das Ausfallrisiko durch einen Austausch innerhalb eines statistischen Sicherheitsintervalls, nehmen dafür aber in Kauf, dass auch technisch einwandfreie Bauteile frühzeitig ausgetauscht werden [Sc10]. Datengetriebene prädiktive Strategien versuchen die Ressourceneffizienz zu erhöhen, indem Sensortechnologien zu Monitoring und Prognose der potenziellen Restlebenszeit eingesetzt werden, gehen aber dafür mit einem Bedarf an zusätzlicher Informationstechnologie einher [Wi19]. Ist ein Störvorfall eingetreten, kann entweder kurzfristig ein lagerndes Ersatzteil zum Kunden geschickt werden oder bspw. mittels additiver Fertigungsverfahren eine kostenintensivere, aber nachhaltigere Reparatur vorgenommen werden [Br19]. Technische Lösungen wie bspw. Service-Support-Systeme, bei denen mittels aktueller Endgeräte Kunden befähigt werden, Reparaturen selbstständig auszuführen, bieten auch eine Alternative zur Serviceerbringung durch einen Techniker [Ka19a, Ni16]. Während der Service schneller und durch die eingesparte Reise auch ressourceneffizienter ausgeführt werden kann, müssen möglicherweise Abstriche in der Servicequalität in Kauf genommen werden.

Die Abwägung über den Einsatz der genannten Beispiele ist keineswegs eindeutig oder sogar stabil zwischen einzelnen Anwendungsfällen übertragbar: Denn während die Minimierung der „time-to-fix“ (der Zeit bis zur vollständigen Entstörung) verfolgt wird, wie es bspw. durch das Lean Management [MS12] vorgeschlagen wird, zielen viele Service-Innovationen auch auf weitere Aspekte ab, die für Kunden mit weniger kritischen Produktionsbedingungen ebenfalls im Vordergrund stehen können. Neben der Entstörungsgeschwindigkeit gewinnen Erbringungskosten und -qualität, aber auch die Erfüllung regulatorischer Rahmenbedingungen und Nachhaltigkeitsziele [MM13] kundenseitig an Bedeutung.

Die Akteure in Wertschöpfungssysteme stehen vor der Herausforderung, ihr Leistungsangebot auf Basis verfügbarer Service-Innovationen anzupassen und kundenzentriert zuzuschneiden. Hieraus ergeben sich zwei zentrale Fragen, die für die Entwicklung von solchen „Tailored Services“ beantwortet werden müssen: (1) Was kann der Service-Erbringer mit den im Wertschöpfungssystem verfügbaren Ressourcen überhaupt sinnvoll lösen? (2) Welche Varianten adressieren bestmöglich die individuellen Prioritäten des Kunden? Arbeiten zur Modularisierung und Konfiguration von Dienstleistungen greifen das Problem auf und bieten erste methodische Lösungen [GBT20]. Gleichzeitig verbleibt die Herausforderung, Tailored Services auch tatsächlich während der Ausführung von Instandsetzungsmaßnahmen kundenspezifisch zu konfigurieren. Wir stellen daher die Forschungsfrage:

FF: Wie können Serviceangebote automatisch an kundenindividuelle Bedürfnisse angepasst werden?

Der vorliegende Beitrag liefert einen Mechanismus für die Konfiguration von After-Sales-Services anhand der verfügbaren Ressourcen und der individuellen Wertvorstellung des Kunden. Wir belegen hierfür zunächst das Variantenproblem bei Instandsetzungsmaßnahmen empirisch anhand von Experteninterviews und modellieren einen Lösungsraum. An-

schließend führen wir die Konfiguration mit einem zweistufigen Mechanismus durch, indem zunächst Aktivitäten eliminiert werden, die aus Ressourcenperspektive nicht durchgeführt werden können. Anschließend werden die verbleibenden Varianten im Lösungsraum anhand kundenindividueller Wertvorstellungen priorisiert. Abschließend demonstrieren wir den Mechanismus an einem Beispiel aus der Landtechnik.

2 Modularisierte Services im Maschinen- und Anlagenbau

Der Lebenszyklus von Maschinen und Anlagen umfasst die Phasen der Entstehung, Vermarktung und die After-Sales-Phase [Ha99]. Mit der Bündelung von Produkten und begleitenden Dienstleistungen erweitert sich dieser Lebenszyklus über die materiellen Komponenten hinaus [Bl10]. Die After-Sales-Phase umfasst den längsten Teil des Lebenszyklus solcher sogenannten hybriden Leistungsbündel und betrifft im Maschinenbau neben dem Vertrieb von Ersatzteilen auch Instandhaltungsmaßnahmen wie Wartungen, Instandsetzungen oder Verbesserungen [Sc05]. Neben ihrer Funktion als Einnahmequelle dienen After-Sales-Services als Differenzierungsmöglichkeit gegenüber dem Kunden [CAA06]. Die Erbringung dieser Services ist mit der Herausforderung verbunden, Bauteile, Servicetechniker und Equipment oftmals über weltweit verteilte Standorte trotz eingeschränkter Prognosefähigkeit kurzfristig bereitzustellen [GBT20].

Um die notwendige Flexibilität gegenüber dynamischen Kundenbedürfnissen zu erreichen, bietet das Forschungsfeld des Service Engineering bereits methodische Antworten zur Modularisierung. Dabei werden in der Design-Phase Elementarservices entwickelt, die über standardisierte Schnittstellen miteinander rekombiniert werden können [DB13]. Damit kann sowohl das Serviceangebot individualisiert, bei ausreichender Skalierung aber auch die Kooperation verschiedener Akteure im gesamten Wertschöpfungssystem erleichtert werden [PU08]. In der Praxis erfolgt die Konfiguration überwiegend einmalig in Form von Service Level Agreements, sodass im Servicefall das Leistungsangebot bereits vordefiniert ist. Die Komplexität einer modularen Servicelandschaft wird hierbei durch Plattformarchitekturen adressiert, die die kundenindividuelle Konfiguration ermöglichen und diese mit operationalen Ressourcen verknüpfen [Ka19b].

Das anforderungsspezifische Tailoring auf Prozessebene wird zumeist in Bezug auf Prozesse als konkrete Handlungsanleitung untersucht. [DRS13] haben eine quantitative Untersuchung durchgeführt, die die Implikationen dieser Ansätze bei der Modellierung und Ausführung aufzeigt. Bei der Konfiguration wird auf Basis eines Referenzmodells eine Prozessvariante erzeugt, die nur notwendige Elemente enthält. Demgegenüber muss bei der Adaption kein Referenzmodell vorliegen, da Prozessvarianten auf Basis einer Menge an Operationen (bspw. Hinzufügen oder Löschen von Aktivitäten) erstellt werden. Eine Übersicht über mögliche Runtime-Operationen wie sie bspw. für Assistenzsysteme genutzt werden finden sich bei [WRR08]. In beiden Fällen wird vorausgesetzt, dass der bestehende Lösungsraum bekannt ist, sodass eine kundenindividuelle Anpassung zur Laufzeit vorgenommen werden kann.

3 Methodischer Ansatz

Die vorgestellten Beispiele aus der Literatur weisen darauf hin, dass technische Innovationen zur Variantenbildung beitragen. Um diese Beobachtung empirisch abzusichern, wurden Experteninterviews durchgeführt, die eine etablierte Methode in der Wirtschaftsinformatik sind, um qualitatives Wissen zu explizieren [MN06]. In elf semi-strukturierten Interviews wurden Experten per Videotelefonat zu Vorgehen, Herausforderungen und Lösungsansätzen im Servicefall befragt. Die Experten verfügen über mehrjährige Berufserfahrung (>5 Jahre) und sind in leitenden Positionen tätig (vgl. Tab. 1). Die Auswertung erfolgte durch Transkription und qualitative Inhaltsanalyse nach [Ma10] voneinander unabhängig durch zwei Forscher. Bei der Zusammenführung der paraphrasierten Ergebnisse wurde zunächst ein generischer Serviceprozess unter Berücksichtigung aller genannten Varianten und Ausprägungen entwickelt.

Nr.	Position	Maschinenbau-Zweig	Anzahl Ma.	Dauer
1	Geschäftsfeldentwicklung Service	Verpackungs-maschinen	> 3000	38:02
2	Geschäftsfeldstrategie Ersatzteilmanagement	Werkzeugmaschinen	> 6500	48:39
3	Abteilungsleitung Service	Werkzeugmaschinen	> 6500	14:41
4	Abteilungsleitung techn. Kundendienst	Landmaschinen	> 5000	24:32
5	Abteilungsleitung Service	Landmaschinen	> 2500	32:13
6	Vertriebsleitung Additive Fertigung	Werkzeugmaschinen	> 14.000	30:06
7	Leitung Werksinstandhaltung	Turbinenbau	> 5000	38:41
8	Geschäftsfeldentwicklung Service	Werkzeugmaschinen	> 1000	31:13
9	Leitung Global Services & Solutions	Werkzeugmaschinen	> 400	26:31
10	Projektleitung Additive Fertigung	Werkzeugmaschinen	> 35.000	69:53
11	Abteilungsleitung techn. Kundendienst	Landmaschinen	> 12.000	33:58

Tab. 1: Übersicht über Experteninterviews

Um das Modell als Konfigurationsgrundlage nutzen zu können, greifen wir auf Ansätze aus der MC-Methodologie zurück, die in der Konstruktion von Produkten als Werkzeuge bereits etabliert sind [FSB12]. Der Mechanismus verfolgt eine für die Produktkonfiguration typische, modellbasierte Perspektive [AV08]. Das generische Prozessmodell wird hierfür als Lösungsraum verstanden, der die Menge aller Prozessvarianten beinhaltet

[GR17]. Im ersten Schritt erfolgt die Ressourcen-basierte Anpassung des Lösungsraumes, wobei wir uns am von [HJ91] vorgestellten Paradigma orientieren. Demnach ist eine Konfiguration im Wertschöpfungssystem nur dann möglich, wenn die in der Umgebung angebotenen Ressourcen die Bedarfe des Systems (in unserem Fall Instandhaltungsprozess) decken. Im zweiten Schritt werden die Varianten hinsichtlich ihres Kundennutzens bewertet, wobei das Elements-of-Value-Modell (EoV) nach [ASB16] genutzt wird, um unterschiedliche Wertbeiträge sichtbar zu machen. Die Autoren unterscheiden im grundlegenden EoV-Modell funktionale Werte wie bspw. Zeitersparnis und Qualität, emotionale Werte, lebensverändernde Werte und sozialen Einfluss. Weil der Maschinen- und Anlagenbau typischerweise kein Consumer-Markt ist, wird hier die später adaptierte Form für B2B-Umgebungen genutzt [ASC18]. Nach Abschluss der Lösungsraummodellierung wird das ST als Constraint-Optimization-Problem [RN21] formuliert, wobei die Verfügbarkeit von Ressourcen im Wertschöpfungsnetzwerk zur Anpassung des Lösungsraumes und die individuelle Zielfunktion der Kundenprioritäten zur Auswahl der optimalen Alternative genutzt werden.

4 Lösungsraum-Modellierung

Die Lösungsraum-Modellierung erfasst Prozessvarianten, die in den Interviews beschrieben wurden und zeigt die Notwendigkeit der kundenindividuellen Konfiguration. Das Kapitel beschreibt so außerdem einen pragmatischen Ansatz zur Entwicklung von Prozessvarianten.

4.1 Integration von Varianten des Instandsetzungsprozesses

Alle Experteninterviews initiieren den Instandsetzungsprozess mit der Meldung eines Fehlers durch den Kunden, auf die vier generische Schritte folgen: (1) die Erfassung aller prozessrelevanten Informationen, (2) die Diagnose und Bewertung des Servicevorfalls, (3) sofern notwendig die Bereitstellung von Ersatzteilen sowie die (4) Durchführung von notwendigen Serviceaktivitäten. Obgleich der Detailgrad der beschrieben Abläufe sowie die angebotenen Lösungsstrategien zwischen den Expertenbeschreibungen variieren, wird einheitlich beschrieben, dass sich der Prozessablauf im Einzelfall erst zur Laufzeit ergibt. Dieser Umstand stellt eine zentrale Herausforderung bei der Bearbeitung von nicht-linearen, dynamischen Prozessen dar, wie sie bspw. bei Diagnose- oder Reparaturfällen auftreten [Me17].

Bei der Aufnahme von Störmeldungen setzt ein Großteil der Unternehmen auf Servicehotlines mit einem oder mehreren Support-Levels. Dabei werden relevante Daten, wie bspw. der Maschinentyp, das Fehlerbild und der Kunde erfasst. Hierauf aufbauend wird eine Fehlerdiagnose gestellt und die Kritikalität des Vorfalls bewertet, um möglichen Lösungswege zu definieren. Dabei ist zu beachten, dass ein Problem mehrere Ursachen haben kann. Die Analyse der zugrundeliegenden Ursache-Wirkungs-Zusammenhänge erfordert

oft spezialisiertes Know-how, wobei auch Sensorik zur Erfassung von Maschinenzuständen und Prognose von Ausfällen eingesetzt werden kann. Zwei der befragten Experten beschreiben darüber hinaus mit digitalen Maschinenakten sowie Serviceplattformen zentrale Informationssysteme, die Kunden und Herstellern kontinuierlich einen Überblick über die in Betrieb befindlichen Maschinen sowie deren Servicehistorie verschaffen. Der beschriebenen Prozessabschnitt ist in Abb. 2 als BPMN-Modell dargestellt.

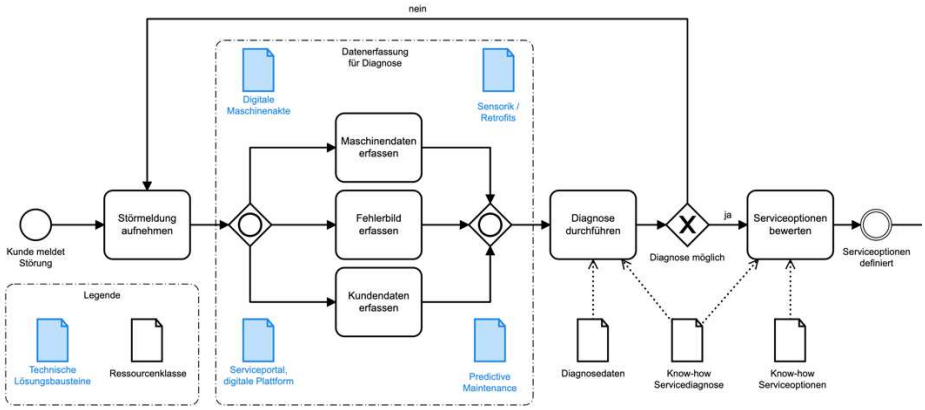


Abb. 2: BPMN-Prozessabschnitt Datenerfassung und Diagnose

Sobald die Serviceoptionen definiert wurden, werden mit der Planung der Ersatzteilbereitstellung und der Reparatur zwei Subprozesse angestoßen, die je nach Unternehmen entweder parallel oder sequenziell bearbeitet werden. Ist die Störung mit einem Ersatzteilbedarf verbunden, muss zunächst das korrekte Bauteil identifiziert werden. In zeitkritischen Fällen werden dem Kunden mehrere Ersatzteile bereitgestellt, um weitere Bauteile als Ursache ausschließen zu können ohne zusätzliche Lieferzeiten zu erzeugen. Je nach Unternehmensstruktur können Bauteile in einem oder mehreren Hauptlagern, in lokalen Hubs, bei Servicepartnern oder dem Kunden gelagert sein. Eine weitere Variante ist die Aufarbeitung beschädigter Teile, die bisher aber überwiegend unabhängig vom konkreten Servicevorfall erfolgt. Gleichzeitig wird aktuell an Lösungen geforscht auch situativ, bspw. mittels additiver Fertigungsverfahren, Bauteile aufzuarbeiten, um den Ressourceneinsatz effizienter zu gestalten. Falls das Bauteil nicht lagernd ist, kann es meist mit hoher Lieferzeit über einen Lieferanten bezogen werden. Alle Experten beschreiben, dass sie anstreben, Ersatzteile in ausreichenden Mengen zu bevorraten, auch wenn Lagerkapazitäten und -kosten limitierende Faktoren darstellen (vgl. Abb. 3).

Nach der Klärung des Ersatzteilbedarfes werden die Serviceaktivitäten geplant. Ein Großteil der Störfälle (ca. 50-80%) kann bspw. bei Softwareproblemen ohne Arbeiten vor Ort behoben werden. Darüber hinaus bestehen drei Varianten: (1) Bei Full-Services wird ein geschulter Techniker geschickt und führt die Arbeiten am Objekt durch. (2) Bei koopera-

tiven Services wird der Kunde per Telefon oder mittels Assistenzsystem durch den Hersteller durch die Prozessschritte geführt. (3) Bei Self-Services führt Kunde den Service ohne direkten Anbieterkontakt selbstständig aus.

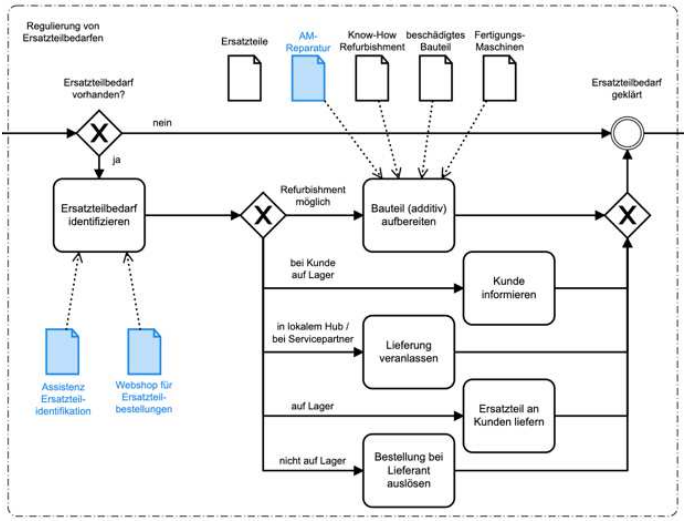


Abb. 3: Prozessabschnitt Ersatzteilbereitstellung

Während mit der Serviceerbringung durch den Kunden Zeit- und Ressourceneinsparungen (u. a. durch Wegfall der Anreise) entstehen, steigt das Risiko, dass der Prozess nicht in hinreichender Qualität durchgeführt wird. Die Entscheidung für eine der Varianten wird in Abstimmung zwischen Anbieter und Kunde getroffen und kann neben der Qualifizierung des Kunden von der Kritikalität des Störfalls, den verfügbaren Ressourcen sowie bestehenden Service Level Agreements abhängen. Gelingt die Entstörung wird der Servicevorfall dokumentiert und abgeschlossen (vgl. Abb. 4).

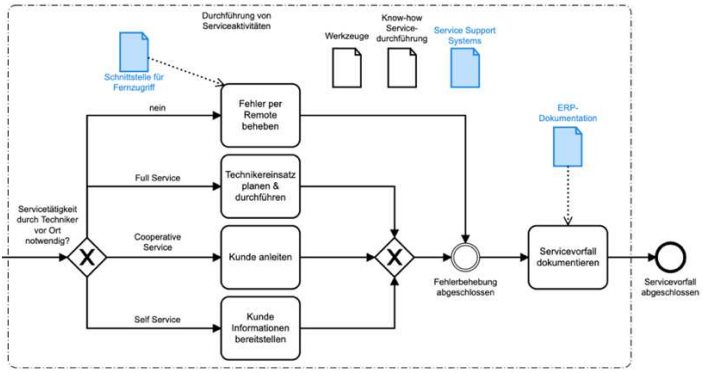


Abb. 4: Prozessabschnitt Serviceaktivitäten

4.2 Gewichtung des Instandsetzungsprozesses mit Elements-of-Value

Neben kurzen Reaktionszeiten beschreiben die Experten weitere Kriterien, die Servicequalität im Maschinenbau definieren. Um den Wert einzelner Prozessvarianten bei Instandsetzungsprozessen zu beurteilen und somit die Leistung kundenindividuell zu konfigurieren, wurden im Zuge der Interviews im Maschinen- und Anlagenbau relevante Zielgrößen identifiziert und einem oder mehreren B2B-EoVs zugeordnet. So beschreibt beispielsweise ein Experte, dass Kunden großen Wert auf die „unmittelbare Erreichbarkeit der Servicehotline“ legen. Hiermit korrespondieren die EoVs *Zeiteinsparung* und *Reaktionsfähigkeit* des Herstellers.

Tab. 2 zeigt die relative Häufigkeit der einzelnen EoVs. Dabei ist auffällig, dass vor allem Zeit, Kosten und Qualität der Leistung für die Kunden ausschlaggebend sind. Darüber hinaus sind der Aufwand und das Eliminieren von Störfaktoren, die mit der Serviceleistung einhergehen, sekundäre Ziele des Prozesses. Mit Verfügbarkeit der Leistung, Expertise des Anbieters sowie Flexibilität in der Ausgestaltung des Service werden auch weitere Anforderungen gestellt, die nur indirekt mit dem konkreten Störfall assoziiert sind.

#	Element of Value	Relative Häufigkeit	#	Element of Value	Relative Häufigkeit
1	Zeitersparnis	100%	8	Expertise	91%
2	Reaktionsfähigkeit	100%	9	Risikoreduktion	91%
3	Kostenreduktion	100%	10	Flexibilität	91%
4	Produktqualität	100%	11	Regul. Compliance	82%
5	Aufwandsreduktion	91%	12	Konfigurierbarkeit	64%
6	Problemreduktion	91%	13	Innovation	55%
7	Verfügbarkeit	91%	14	Soziale Verantwort.	45%

Tab. 2: B2B-Elements of Value im Maschinen- und Anlagenbau

5 Konfiguration von Tailored Services

Um unsere Forschungsfrage zu beantworten, entwickeln wir eine zweistufige Methode zur Konfiguration des Lösungsraums. Hierfür werden zunächst Aktivitäten eliminiert, die durch im Wertschöpfungsnetzwerk nicht erbracht werden können. Anschließend erfolgt die Priorisierung verbleibender Lösungsalternativen anhand ihrer EoV-Gewichtung.

5.1 Einschränkung des Lösungsraumes durch Ressourcenbedarfe

Mögliche Prozessvarianten im Instandsetzungs-Lösungsraum werden durch die situative Ressourcenverfügbarkeit bedingt. Im individuellen Servicevorfall muss daher das Res-

sourcenangebot im Wertschöpfungsnetzwerk (d. h. beim Kunden, beim Anbieter und etwaigen Kooperationspartnern) mit dem Ressourcenbedarf der Prozessvariante abgeglichen werden [vgl. PK21]. Alle Varianten, die für den Einzelfall nicht durchgeführt werden können, weil bspw. ein Ersatzteil nicht beim Kunden lagert oder Werkzeuge für einen Self-Service nicht beim Kunden zur Verfügung stehen, werden bei Entscheidungssituationen im Prozess nicht betrachtet, indem die entsprechenden Pfade für die weitere Konfiguration eliminiert werden. Für modulare, technische Systeme wurde mit der Ressourcenbasierten Konfiguration bereits eine adäquate MC-Methode entwickelt [HJ91], die analog für die Problemstellung des ST angewandt werden kann.

5.2 Auswahl wertmaximierender Prozessvarianten

Um den eingeschränkten Lösungsraum bestmöglich auf kundenindividuelle Bedürfnisse auszurichten, ist die Bestimmung einer optimalen Prozessvariante notwendig. Durch die EoV-Bewertung kann eine Pfadoptimierung im Lösungsraum durchgeführt werden. Damit folgen wir einem Graph-basierten Ansatz, wie er von [No20] bereits auf Prozessgraphen mit einzelnen Kriterien anwendet wurde. Hierzu müssen alternative Lösungsstrategien wie bspw. Versand aus dem Lager vs. Bestellung beim Lieferanten gemäß der EoVs miteinander verglichen und bewertet werden. Der entstehende Vektor *v_i* beschreibt die Bewertung jeder Alternative mit den 14 identifizierten EoVs im Vergleich zu allen alternativen Lösungsstrategien im vorliegenden Prozessschritt, sodass bspw. Lagerversand mit einem Wert von *EoV_Zeitersparnis = 4* doppelt so schnell ist wie der Versand durch den Lieferanten mit einem Wert von *EoV_Zeitersparnis = 2*. Ist ein Knoten nicht bewertet, bspw. weil keine Alternativen zur Verfügung stehen, wird ein Nullvektor eingefügt.

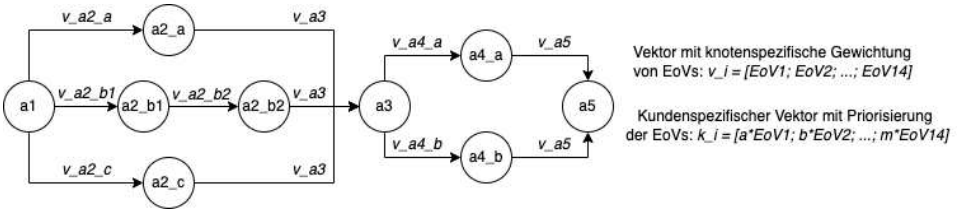


Abb. 5: Lösungsraum für Konfigurationsproblem als gerichteter Graph

Im Falle der individuellen Optimierung wird für den Kunden ein Präferenzvektor *k_i* erstellt, der für jedes der EoVs einen individuellen Gewichtungsfaktor sowie die Relevanz eines EoVs als binärem Wert enthält (vgl. Abb. 5). Die Gewichtung vermindert den Effekt gegensätzlicher Präferenzen (bspw. Zeitersparnis und Kosten). Anschließend wird der bewertete Prozess durch eine Multiplikation aller Kantengewichte mit dem Präferenzvektor instanziiert, sodass ein kundenindividueller, gewichteter Prozessgraph entsteht. Die Identifikation des optimalen Pfades und damit die Auswahl von Lösungsalternativen für die Prozessabschnitte wird als *multiple-criteria longest path problem* beschrieben. Da die

hierfür entwickelten Suchalgorithmen auf Kantengewichten basieren, werden die Gewichte im Prozessgraph von den Knoten auf die vorgelagerten Kanten vorgezogen, um die Komplexität des Graphen nicht durch weitere Knoten zu erhöhen [vgl. No20].

5.3 Demonstration des zweistufigen Service-Tailoring-Mechanismus

Insbesondere bei saisonal genutzten Maschinen wie der Landwirtschaft variieren Kundenprioritäten im Jahresverlauf. Durch einen Experten aus dem Landmaschinenbau wurde beschrieben, dass die Kunden zur Erntezeit Zeiteinsparungen bei der Reparatur priorisieren, aber in den Wintermonaten den Fokus auf die Qualität der Reparatur legen, um während der Saison Ausfälle zu vermeiden. Anhand dieses Beispiels lässt sich der vorgestellte Konfigurationsmechanismus demonstrieren: In Abb. 6 ist ein vereinfachter Ausschnitt des generischen Serviceprozesses für zwei Kunden mit identischem Schadensfall dargestellt. Eine Refurbishment des defekten Bauteils kann mangels Know-How nicht angeboten werden und schränkt den Lösungsraum ein. Alle Kundenoptionen wurden exemplarisch hinsichtlich Zeit- und Kostenersparnis sowie Qualität bewertet. Die Reparatur des defekten Bauteils ist im Vergleich zum Einkauf eines Lagerteils teurer, aber schneller verfügbar als eine Lieferung aus dem Zentrallager. Der Full-Service durch einen ausgebildeten Techniker zeichnet sich durch hohe Qualität aus, dauert aber länger als eine Reparatur durch den Kunden im Self-Service. Je nach Kundenanforderungen ergeben sich unterschiedliche optimale Pfade *(bspw. v_a2b (k1))*. Ist die betroffene Maschine in einen zeitkritischen Auftrag eingebunden, würde der Fokus zunächst auf der Zeiteinsparung liegen *(K1)*. In diesem Fall bietet sich die Nutzung des Lagerteiles sowie der Self-Service an. Ist die Maschine wie bspw. bei Landmaschinen üblich über den Winter nicht ausgelastet, liegt der Fokus des Instandsetzungsprozesses auf der Qualität des Austausches *(K2)*, da die Verfügbarkeit in der Saison sichergestellt werden soll. In diesem Falle liegt die optimale Lösung in der Reparatur durch einen Servicetechniker im Full-Service.

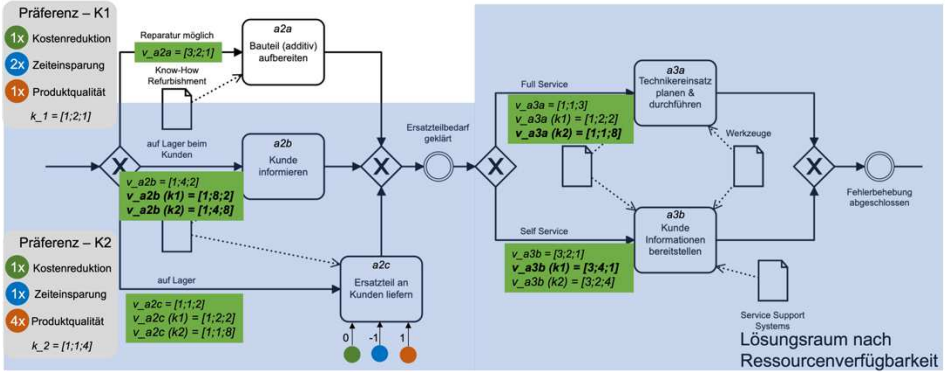


Abb. 6: Tailored-Service-Anwendung im Landmaschinenbau

6 Diskussion & Fazit

Das vorliegende Papier baut auf der Beobachtung auf, dass der Maschinen- und Anlagenbau eine Reihe von technischen Innovationen im Aftermarket einführt und betreibt, die zur Optimierung des Maschinenbetriebs beitragen sollen und unterschiedliche Ziele, wie beispielsweise die Steigerung der Ausfallsicherheit, der Ressourceneffizienz oder der Instandsetzungszeit verfolgen. Wir beschäftigten uns deshalb mit der Frage, wie die entstehenden Serviceangebote automatisiert auf kundenindividuelle Bedürfnisse zugeschnitten werden können. Wir belegen hierfür die zunehmende Variantenvielfalt im Aftermarket anhand von Experteninterviews und schlagen einen zweistufigen der MC entlehnten Lösungsmechanismus vor. Durch die Demonstration in der Landtechnik zeigen wir den konkreten, praktischen Nutzen des Service Tailorings. Weil wir auf Ebene der allgemeinen Servicemodellierung ansetzen und nicht auf branchenspezifische Modelle zurückgreifen, erwarten wir eine Übertragbarkeit des Mechanismus auch auf andere Dienstleistungsfelder. Allerdings ist zu berücksichtigen, dass der Konfigurationsmechanismus voraussetzt, dass Varianten aus EoV-Perspektive quantifiziert werden können. Hierin liegt eine Herausforderung für die Praxis, Kundenbedürfnisse detailliert verstehen und bewerten zu können. Aus theoretischer Perspektive trägt das ST eine Kernfunktion für die Entwicklung kontextadaptiver „Smart Services“ bei. Diese entfalten vor allem dann ihren Mehrwert, wenn die Auswahl einer optimalen Strategie aufgrund von Reaktionszeit oder Komplexität manuell nicht mehr beherrschbar ist. Ein sinnvoller nächster Schritt erscheint uns in diesem Sinne die Erweiterung des ST um Kontextwahrnehmung, wie sie beispielsweise von [Ka19c] vorgeschlagen werden. ST ist dabei nicht als einmalig zu entwickelnde, abgeschlossene Funktion zu betrachten, sondern erfordert kontinuierliche Weiterentwicklung [vgl. Th17]. Technologische Innovationen werden auch in Zukunft neue Servicevarianten eröffnen und so Lösungsräume schaffen, in denen Kundenbedürfnisse noch individueller adressiert werden können.

Literaturverzeichnis

[ACS18] Almquist, E., Cleghorn, J., & Sherer, L.: The B2B Elements of Value. Harvard Business Review, 96/3, S. 72–81, 2018.

[ASB16] Almquist, E., Senior, J., & Bloch, N.: The Elements of Value. Harvard Business Review, 94/9, S. 47–53, 2016.

[AV08] Aldanondo, M., & Vareilles, E.: Configuration for mass customization: How to extend product configuration towards requirements and process configuration. Journal of Intelligent Manufacturing, 19/5, S. 521–535, 2008.

[Bl10] Blinn, N., Nüttgens, M., Schlicker, M., Thomas, O., & Walter, P.: Lebenszyklusmodelle hybrider Wertschöpfung: Modellimplikationen und Fallstudie. In (O. Thomas, P. Loos, & M. Nüttgens, Hrsg.), Hybride Wertschöpfung, S. 130–143, Springer Berlin Heidelberg, 2010.

[Br19] Brinker, J., Gembarski, P. C., Hagen, S., & Thomas, O.: Anwendungspotenziale von Additive Repair und Refurbishment für Service-orientierte Geschäftsmodelle. In (R. Lachmayer, K. Rettschlag, & S. Kaierle, Hrsg.), Konstruktion für die Additive Fertigung 2019, S. 43–54, Springer Berlin Heidelberg, 2020.

[CAA06] Cohen, M. A., Agrawal, N., & Agrawal, V: Winning in the Aftermarket. Harvard Business Review, 84/5, S. 129–138, 2006.

[DB13] Dörbecker, R., & Böhmann, T.: The concept and effects of service modularity–A literature review. 46th Hawaii International Conference on System Sciences, S. 1357–1366, 2013.

[DRS14] Döhring, M., Reijers, H. A., & Smirnov, S.: Configuration vs. adaptation for business process variant maintenance: An empirical study. Information Systems, 39, S. 108–133, 2014.

[FSB12] Fogliatto, F. S., da Silveira, G. J. C., & Borenstein, D.: The mass customization decade: An updated review of the literature. International Journal of Production Economics, 138/1, S. 14–25, 2012.

[GBT20] Gembarski, P. C., Brinker, J., & Thomas, O.: An extended key competence framework to mass customize after-sales services. Proceedings of the 9th International Conference on Mass Customization and Personalization-Community of Europe (MCP-CE 2020), S. 67–76, 2020.

[GR17] Gembarski, P. C., & Lachmayer, R.: Mass Customization und Product-Service-Systems: Vergleich der Unternehmenstypen und der Entwicklungsumgebungen. In (O. Thomas, M. Nüttgens, & M. Fellmann, Hrsg.), Smart Service Engineering, S. 214–232, Springer Fachmedien Wiesbaden, 2017.

[Ha99] Harms, V: Kundendienstmanagement: Dienstleistung, Kundendienst, Servicestrukturen und Serviceprodukte; Aufgabenbereiche und Organisation des Kundendienstes. Neue Wirtschaftsbriefe, Herne, Berlin 1999.

[HJ91] Heinrich, M. & Jüngst, E.W.: A resource-based paradigm for the configuring of technical systems from modular components. Proc. CAIA'91, S.257–264, 1991.

[JCG18] Jacobides, M. G., Cennamo, C., & Gawer, A.: Towards a theory of ecosystems. Strategic Management Journal, 39/8, S. 2255–2276, 2018.

[Ka19a] Kammler, F., Brinker, J., Vogel, J., Hmaid, T., & Thomas, O.: How Do We Support Technical Tasks in the Age of Augmented Reality? Some Evidence from Prototyping in Mechanical Engineering. ICIS 2019 Proceedings. International Conference on Information Systems (ICIS), Munich, Germany, 2019.

[Ka19b] Kammler, F., Acik, E., Brinker, J., Thomas, O. & Nüttgens, M.: Digitale Plattformstrategien für hybride Wertschöpfungssysteme: Exploration und Wirkungsanalyse im Maschinen-und Anlagenbau. In (V. Stich, J. H. Schumann, D. Beverungen, G. Gudergan, P. Jussen, Hrsg.), Digitale Dienstleistungsinnovationen, S. 579–599. Springer Vieweg, Berlin, Heidelberg, 2019.

[Ka19c] Kammler, F., Hagen, S., Brinker, J. & Thomas, O.: Leveraging the Value of Data-Driven Service Systems in Manufacturing: A Graph-Based Approach, Proceedings of the 27th European Conference on Information Systems, Stockholm & Uppsala, Sweden, S. 1–15, 2019.

[Ma10] Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken (11., aktualisierte und überarb. Aufl). Beltz, 2010.

[Me17] Metzger, D., Niemöller, C., Berkemeier, L., Brenning, L., & Thomas, O.: Vom Techniker zum Modellierer – Konzeption und Entwicklung eines Smart Glasses Systems zur Laufzeitmodellierung von Dienstleistungsprozessen. In (O. Thomas, M. Nüttgens, & M. Fellmann, Hrsg.), Smart Service Engineering, S. 193–213. Springer Fachmedien Wiesbaden, 2017.

[MM13] Martínez-Jurado, P. J., & Moyano-Fuentes, J.: Lean Management, Supply Chain Management and Sustainability: A Literature Review. Journal of Cleaner Production, 85, S. 134–150, 2014.

[MN06] Myers, M. D., & Newman, M.: The qualitative interview in IS research: Examining the craft. Information and Organization, 17/1, S. 2–26, 2006.

[MS12] Moyano-Fuentes, J., & Sacristán-Díaz, M.: Learning on lean: A review of thinking and research. International Journal of Operations & Production Management, 32/5, S. 551–582, 2012.

[Ni16] Niemöller, C., Metzger, D., Fellmann, M., Özcan, D., & Thomas, O.: Shaping the future of mobile service support systems-ex-ante evaluation of smart glasses in technical customer service processes. In (H. C. Mayr & M. Pinzger, Hrsg.), INFORMATIK 2016, Lecture Notes in Informatics (LNI), S. 753–767, Gesellschaft für Informatik (GI), 2016.

[No20] Nordemann, F., Tönjes, R., Pulvermüller, E., & Tapken, H.: Graph-based multi-criteria optimization for business processes. International Symposium on Business Modeling and Software Design, S. 69–83. Springer, Cham, 2020.

[PU21] Gembarski, P. C., & Kammler, F.: Mass Customizing for Circular and Sharing Economies: A Resource-Based View on Outside of the Box Scenarios. Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems, S. 1039–1046. Springer, Cham, 2021.

[PU08] Pekkarinen, S., & Ulkuniemi, P.: Modularity in developing business services by platform approach. The International Journal of Logistics Management, 19/1, S. 84–103, 2008.

[RN21] Russell, S. J., & Norvig, P.: Artificial intelligence: A modern approach (Fourth edition). Pearson, 2021.

[Sc05] Schild, U.: Lebenszyklusrechnung und lebenszyklusbezogenes Zielkostemanagement. Deutscher Universitätsverlag Wiesbaden 2005.

[Sc10] Schenk, M.: Instandhaltung technischer Systeme. Springer Berlin Heidelberg, 2010.

[Th17] Thomas, O., Varwig, A., Kammler, F., Zobel, B. & Fuchs, A.: DevOps: IT-Entwicklung im Industrie 4.0-Zeitalter. HMD Praxis der Wirtschaftsinformatik, 54/2, S.178–188, 2017.

[Th21] The True Cost of Downtime – How much do leading manufacturers lose through inefficient maintenance?, https://www.senseye.io/blog/the-true-cost-of-downtime, abgerufen am: 01.05.2021.

[Wi19] Wiedemann, N., Kammler, F., Varwig, A., & Thomas, O.: Towards a Framework for Predictive Maintenance Strategies in Mechanical Engineering–A Method-Oriented Literature Analysis. In (T. Ludwig & V. Pipek, Hrsg.): 14. Internationale Tagung Wirtschaftsinformatik, 2019.

[WRR08] Weber, B., Reichert, M., & Rinderle-Ma, S.: Change patterns and change support features – Enhancing flexibility in process-aware information systems. Data & Knowledge Engineering, 66/3, S. 438–466, 2008.

A Platform Framework for the Adoption and Operation of ML-based Smart Services in the Data Ecosystem of Smart Living

Henrik Kortum[1,3], Tobias Kohl[2], Dominik Hubertus[3], Oliver Hinz[2] and Oliver Thomas[1]

Abstract: Smart services utilizing machine learning (ML) take a more and more important position in our daily lives. As a result, the need for a large smart living data ecosystem has emerged that links the most diverse areas of life with each other. This ecosystem is characterized by a multitude of different actors, a heterogeneous system, product and service landscape as well as high data protection requirements. To provide real added value and holistic solutions in this tension field, the orchestration of different subservices is necessary, bundling the functionality of individual smart devices and models. For this goal to be achieved, a framework that considers the complex challenges of the ecosystem focusing on the adoption and operation of smart services is required. Here our paper makes a key contribution. Based on requirements from the literature and concrete smart living use cases, we derive a platform framework for this data ecosystem.

Keywords: ML, Smart Services, Smart Living Framework, ML-Adoption, Data Ecosystem.

1 Introduction

Smart services utilizing machine learning (ML) take a more and more important position in our daily lives. They have shown to be pivotal to address societal concerns such as care for elder citizens, by developing and utilizing specialized ML models [ZSBC21]. For the domain of smart living, the applications scenarios of smart services are therefore no longer limited to simple applications in the smart home environment but grow to be more complex and interconnected. The domain is increasingly evolving into a data ecosystem [KRHT22] in which an interplay of different actors, such as organizations and actual individuals, have to share data in networks and use it as a resource [OlLó18]. These actors need to encompass use cases such as assisted living, predictive maintenance and energetic sustainability, while also addressing not only smart home environments but the whole residential context on scale [KoKM21, LoMi21]. While a variety of new technologies and services are being developed [WuVS19], many of them still represent standalone solutions [BHFR19] that cannot be easily integrated into a holistic view of the complex and highly

[1] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH – Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, Germany, {henrik.kortum, oliver.thomas}@dfki.de
[2] Goethe-Universität Frankfurt – Wirtschaftsinformatik und Informationsmanagement, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, Germany , {tokohl, hinz}@wiwi.uni-frankfurt.de
[3] Strategion GmbH, Albert-Einstein-Straße 1, 49076 Osnabrück, Germany, {henrik.kortum, dominik.hubertus}@strategion.de

doi:10.18420/inf2022_32

interconnected data ecosystem that the smart living domain represents. As such for Service Systems Engineering (SSE) broader domain challenges are often not considered. We propose a framework designed to facilitate societal value creation in the smart living domain by focusing on both synchronous and asynchronous smart services, as well as on a service orchestration component that facilitates a secure exchange and processing of data between all market actors, while also providing full agency of the shared data to the data owners. To ground our proposed platform framework, we first analyze different domain use cases in literature as well as from technical service descriptions. Here we motivate the challenges that in turn base the requirements we identify in the smart living data ecosystem. Building on this foundation, our paper aims to provide answers to the following two research questions.

RQ1: What domain challenges exist in the smart living ecosystem and which requirements for the adoption and operation of ML-services can be informed from them?

RQ2: How does a platform framework for the adoption and operation of ML-services have to function to address the identified domain requirements (RQ1) and enable a value creation process in the smart living ecosystem?

In this paper we therefore first ascertain practical requirements and challenges in the smart living data ecosystem and then aim to propose a framework for a platform architecture for ML based smart services addressing the challenges identified. Just like the smart service systems, a platform architecture itself should not be designed monolithically and operated as a centralized system by one Hyperscaler. Rather, the platform components must be conceived on a functional level and designed in a modular way. Not least in order to strengthen the trust of the actors using the ecosystem and to guarantee its openness. Smart service providers face a multitude of domain challenges if they want to offer smart services for the smart living ecosystem. One such challenge is that many proposed high performing ML models for the domain use cases, like Wave Net models for energy disaggregation [JKPG19], are not designed to work in environments with realistic data occurrence [ChHo19]. To address this disparity, the smart service model design can be modified [KoKM21], but this would also require the ecosystem to facilitate the exchange of data and knowledge between customers, who are often real individuals also acting as data owners, and service providers. More problematic however is the forementioned heterogenic nature of most environments in the smart living domain as many ML models are very limited in their transferability. For instance energy consumption model patterns often change between users and appliances [SRLR19, ZGIR12]. The recognition of patterns across multiple apartments will be hindered by these variations in the data. As such models often need to be retrained for specific households or even just for new residents. Accordingly, the training process of ML models represents a central component for the value creation in smart living. As service providers often do not own the data needed to train models and provide meaningful services the secure and robust exchange of data between data owners and service providers plays a critical role for this human centered ecosystems [KRHT22]. For service providers to develop services and ultimately generate societal value the data owners (e.g. residents) must themselves take a strong interest to make their personal data available and take a stake in facilitating a meaningful data allocation. This

in turn however could conflict with their own interests as users, as well as with broader domain requirements such as privacy protection and data security. User acceptance is an important factor for technology adoption in any domain and is bound to performance and effort expectancy beyond the feature sets of services [VeTX16]. Specifically, in the smart living domain as a human centered ecosystems the adoption of services by users and the willingness to share their data is intrinsically bound to their perceived data sovereignty. This fact requires the consideration of general conditions such as data security, transparency and explainability to increase acceptance [RKGE22]. The loss of privacy is one of the greatest fears of residents in smart environments [WuVS19], which is why solutions in the ecosystem must explicitly take the above aspects into account. For the creation of societal value, the fast processing of data by ML-services has a prominent role to play as many IoT devices generate data at short intervals, accordingly the processing must often be carried out in near real-time. Following a complementary ML perspective focused on requirements for both model training and continuous real time data processing is necessary. only to address the technical phases for ML-service training and deploying.

2 Methodology

We rely on the Design Science Research Methodology (DSRM) by [PTRC07] for structuring our approach (see Fig. 1). The DSRM provides a procedure that is based on previous IS literature and adapts it for the DSR Contribution Framework. It starts with identifying the problems and challenges that are fundamental for defining the goals of the artifact in the domain. As suggested in [GrHe13], we incorporate existing knowledge from the domain in form of the analyzed literature and technical service concepts for design and development. Specifically, we apply a multi-method approach [BrHO06] in this work, combining a literature search with a qualitative document analysis. To achieve an overview depicting the specific complex challenges of the domain we start with a non-structured literature review for domain requirements in smart living. We focus on topics such as data exchange between actors, adoption by users, and data ownership in a sophisticated smart living ecosystem. For complementation we further conducted a document analysis following [Bowe09] to identify core concepts related to domain requirements in practice. A total of 38 service concept documents were evaluated, which correspond to the overarching smart living use cases: “assisted living”, “smart kitchen”, “management of smart buildings & predictive maintenance”, “intelligent gatekeeper for building access”, “smart energy management” and “software defined sensors”. Each concept document was created by a team of managers and technicians with several years of professional experience in the smart living domain. The experts cover all relevant sectors of the ecosystem, such as the housing industry, IoT and product manufacturers, ML developers, infrastructure providers and various service companies. Through triangulation, we consolidate both result strands, and derive a requirement set consisting of 21 requirements, thus answering RQ1. The requirements set is the basis for our proposed framework, with individual components contributing to the implementation of individual requirements, answering RQ2.

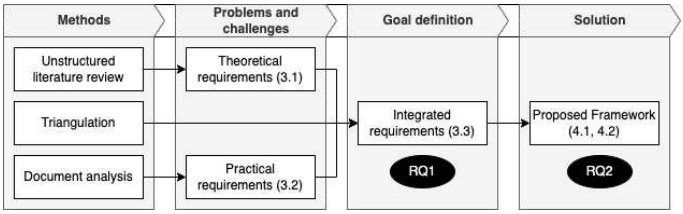


Fig. 1: Visualization of the DSR oriented approach

3 Domain Challenges and Requirements

3.1 Domain specific requirement elicitation in the literature

This section collects and presents domain-specific requirements from the literature. An increasing number of technologies are being developed for the smart living domain [WuVS19]. However, there is a lack of holistic approaches that combine different perspectives such as user requirements, ML specificities, and a distribution of data sources among different actors in the data ecosystem [KGZF20]. User acceptance and the general consideration of human factors in the development of smart living services often fall short [Grun20]. Yet, paradoxically, user-centricity is especially important for the smart living domain, where a very diverse user base, varying greatly in age, gender, ethnicity and level of knowledge, needs to be considered [Grun20]. Taking these aspects into account also has an impact on the technical implementation, e.g. by integrating tests for unwanted bias in the ML-services to avoid discrimination against certain subgroups and thus increase user trust [KFRE22, RKGE22]. One of the biggest obstacles to the acceptance of smart living solutions by users is their fear of losing privacy [WuVS19]. Smart home devices such as smart meters store data on consumption patterns, which can be used to derive sensitive information's such as usage habits or social status [WuVS19]. The resulting demand for ensuring privacy is therefore one of the most central requirements for services in the smart living data ecosystem [CBBF19, SCDM16, WaGM17, WuVS19, ZNYL17]. The application scenarios of the ecosystem extend far beyond simple home automation and also include further areas, such as smart energy management, health or assisted living [NiDe15, SoGB13]. Considering the interplay of multiple data sources like sensor data and energy and water consumption with external data sources, such as weather forecasts or air pollution data [BDMB17], and integrating them into holistic models is therefore an important domain challenge [ZSBC21]. All this information needs to be linked in a meaningful way to generate real value-added services [KGZF20]. While the domain actors demand traceability and integrity of data flows throughout the entire lifecycle of smart living solutions, across different services and providers [KRHT22]. Especially when the decisions derived from sensor data by ML-services may result in serious consequences, the integrity of the data streams must be ensured [CBBF19]. However, such integrated view is hindered by the stand-alone nature of many smart living devices [BHFR19]. Research in

the area of smart living therefore increasingly focuses on creating interoperability of products, data, and services within the ecosystem [BSSK17, MiRF19]. Platform architectures in this environment should therefore consider components supporting the integration of these multi-layered data sources and the communication of individual components [BDMB17]. Depending on the use case, the platform should be able to support both synchronous and asynchronous communication [BDMB17].

Many systems and IoT devices in the smart living environment rely on a combination of edge and cloud computing [WaYL18], which means, especially in the cloud case, that data must be encrypted to protect the privacy of the data owner [ZNYL17]. This privacy requirement competes with the ecosystem's need to share data between different systems and actors, especially when implementing more complex machine learning tasks such as activity recognition. In this context, data sovereignty must always lie with the data owner, who should be able to control the distribution of his data at any time, resulting in specific requirements for identity and access management. Therefore, authorization management and anonymization concepts need to be developed that allow data to be shared among stakeholders while maintaining privacy.

3.2 Domain specific requirements based on use case descriptions

Following the DSR model, we identify domain challenges not only on the basis of the scientific literature but also by examining concrete use cases through document analysis [Bowe09]. To this end, the main findings are briefly summarized in the following and subsequently classified in a category scheme together with the literature requirements in section 3.3. The investigated use case documents cover a wide range of application scenarios, including simple ML-based assistance services for the kitchen, anomaly detection and predictive maintenance for sensor data, activity recognition, intelligent access management, as well as more critical use cases such as an emergency detection. By analyzing the use case documents, we were able to confirm that the smart living data ecosystem is characterized by a multitude of heterogeneous IoT devices that generate large amounts of data in a very short time. Services for maintenance or activity recognition, for example, require real-time evaluation of high-frequency sensor data from different providers. The data needed to realize the various use cases is often distributed among different providers and sensor manufacturers. To allow meaningful integration, the systems must support interoperability of data sources and services. A consequence of this heterogeneity is also the demand often made in the use case documents for a central overview of the services and data sources available in the data ecosystem. This should contain information on how data sources can be integrated and how third-party services can be embedded in the use case. The document analysis also shows that transparency and traceability are important requirements in the context of ML models. Often it is necessary to understand why a model decision was made and on which data it was trained. This is especially important in application scenarios, where the potential negative consequences can be enormous (e.g. for the use case of emergency detection). Such critical use cases must guarantee a professionally tested quality of the models. The quality requirements relate not only to the ML model but

also to the robustness and fault tolerance of the service. The service for emergency control of the building doors, for example, must be always available. The document analysis confirms the strong user-centricity of smart living use cases. As a result, the data processed in most use cases is very sensitive and must be protected, as it allows conclusions to be drawn about the activities of the residents, their preferences, and habits. Many of the use cases described concretize the demand from the literature for measures to preserve the privacy and data governance of the data owner. It is interesting that most of the services are planned as B2B services and address the housing industry, but for their operation, the resident as data owner must still actively contribute his or her data. An example is the service of the intelligent gatekeeper, which automatically controls access to the building for authorized persons. According to the use case description, the service supports, among other access methods, a facial recognition for residents. To protect sensitive biometric data, our framework must support a complex assignment of rights by the data owner and be reliable in terms of data protection and security. In other use cases, such as sensor data analysis, the authentication implications go even deeper, requiring an authentication concept at the level of specific sensor attributes that allows selective authorization by the data owner for specific purposes. Such a permission must be revocable by the data owner at any time. The platform framework must ensure that ML-services and other applications that have used this data are unable to do so from the time of withdrawal and, if required, delete stored data (e.g. in training sets). To ensure that these requirements are adequately taken into account, many of the analyzed services provide for the involvement of domain experts. Further requirements result from the dynamic character of smart living data: As resident behavior changes over time or new people may join a household, several use cases require continuous retraining of the ML models after deployment. As an example, the use case for energy consumption forecasting can be mentioned here, which has to react to changes in the composition of the household and thus to changing consumption patterns. All use case documents that explicitly include ML algorithms require a computationally powerful infrastructure for the initial model training. In addition, some use cases demand that once trained and integrated into the final service, the models must be operable on edge with limited hardware. One example is emergency detection, which uses a multisensory principle to determine whether a person has suffered an accident and has, for example, fallen badly or needs help in some other way. Such model has to be trained on a large amount of multivariate time series data, using considerable computing power, while the final ML-service must be run on wearable and edge devices. Since errors and failures in this, as in many other use cases, can have serious consequences, high requirements are placed on the robustness and failure safety of the system. This also and in particular applies to the processing of data by ML-services. The guarantee of integrity requirements is explicitly demanded, since unwanted changes to the input data can, for example, lead to an emergency not being recognized.

3.3 Triangulation and Requirement Conceptualization

In the following, findings from the literature search are merged with results of the document analysis, using the method of triangulation [BrHO06]. The resulting set will serve as basis for developing a platform concept. Fig. 2 shows the triangulation results. The boxes shortly describe the identified requirements grouped by the following identified categories: Service-related, ML-related, Data-related, System-related and Consumer-related.

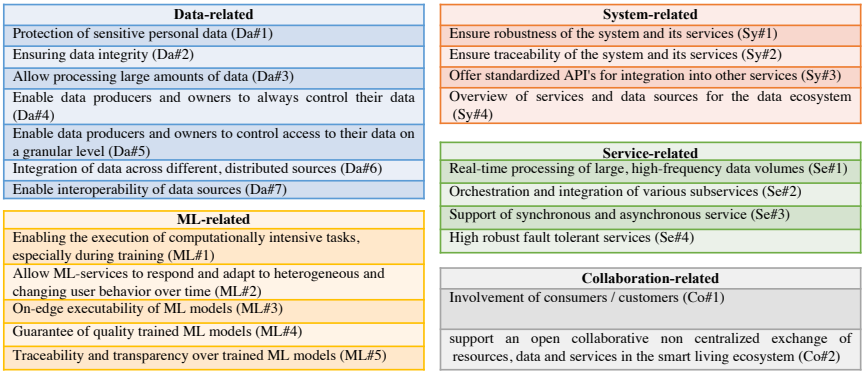


Fig. 2: Conceptualized domain requirements for Smart Living

Each requirement contains a unique ID, which we use in section 4 to create a concrete link between elements of the solution artifact and the individual requirements.

4 Framework for a Smart Living Platform Architecture

4.1 Solution proposal

To address the smart living ecosystem means to address a large variety of service solution providers, both private and public, who are offering services supporting all kinds of use cases in the residential context with each service warranting a different set of requirements for data access and security. It also means addressing the housing industry at scale, which uses a variety of differing IoT hardware devices resulting in a variety of distinct, non-standard sensory data formats and configurations (Da#7). Also, each residential building provides a heterogeneous landscape of IoT devices in form of the home environments of the residents, who also act as supplemental auxiliary data owners in the ecosystem (Da#6, ML#2). For event-based reasoning data from these IoT devices will therefore be structured in an interoperable IoT Web-of-Thing (WoT) Schema (Da#7). Following this approach models can be created to work on an abstracted layer, that offers specific properties for each thing, which are defined and unified between different sensors (ML#2). This way models and services can work on a multitude of heterogeneous devices and sensor configurations. A platform framework for the adoption and operation of smart services in smart living needs to address the requirements and challenges of both sides. On a technical level

it needs to facilitate the secure exchange of data between all actors in the ecosystem (Da#2, Da#6, Da#7) and the reliable orchestration of smart services (Se#2, Sy#1, Sy#2), while offering full agency over the used data to the data owners (Da#4), which are both housing industry and residents, with both also acting as service customers. Based on these requirements the framework needs to operate in form like a two sided market platform, generating societal value through the interactions between two distinct groups of platform actors [JKPS19], in this case data owners and service providers (see Fig. 3).

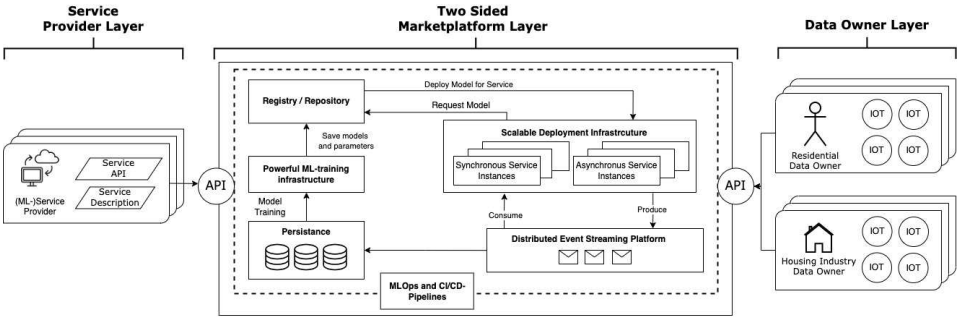


Fig. 3: Proposed platform framework

As a two-sided market platform the ecosystem is fully bound to network externality effects for its growth, i.e. the market adoption by data owners, service providers and service customers [HiOS20, JKPS19]. Therefore, the platform needs to ensure the cooperation and adoption by a wide variety of service providers, both for services targeting housing environments of residents but also those offering services for housing providers targeting the management of the residential context on scale. While a certain level of both service customer and service provider adoption will facilitate the adoption by others [HiOS20] certain technical platform capabilities for scalability (Sy#3) and ease of adoption will be a requirement for consideration. As such the proposed platform supports a scalable deployment infrastructure (Se#2, Sy#1) for all service types (Se#3). However, the primary challenge for service providers remains in getting models from research environments to productive use in services for real housing environments as most models are not highly transferable between environments. We therefore also propose a Machine Learning Operations (MLOps) pipeline to streamline the development (ML#4, ML#5) and rollout process for service providers, which requires the inclusion of a powerful compute environment for model training (ML#1) into the platform ecosystem. While the intrinsic motivation of service providers is profit oriented, we will leave platform business models for different actors, such as the data owner, and services in the smart living ecosystem open for further research and not cover in this paper. For service customers the platform itself will need to act as an enabler for the value creation process, as from their perspective they only interact with the platform not with each individual service provider or actor. As such we propose a platform based smart service store offering and allocating smart services tailored to the knowledge and technological capabilities of each service customer's environment (Sy#4). This is also where both types of service customers meaning residents and housing industry can interact with each other (Da#6), as both can provide access to services and data to

each other. Important for the acceptance here is to provide residents with a granular authorization concept (Da#5) for each data property in their environment not only for each booked service but also for the sensor data to be shared with their housing providers. Such authorization concept has shown to be critical to protect sensitive data and secure the interests of residents (Da#1), as residential personal data can inform housing providers about user preferences and habits indicating social status and cultural identity [WuVS19]. It's not hard to imagine topics such as mold prevention sparking conflict between residents and their housing providers resulting in legal actions or housing contract termination.

Most of the technical challenges for the adoption and operation of ML-services are also defined by the use cases in the smart living domain. IoT devices from home and housing environments generate a large volume of high-frequency data, whose processing for value creation needs to happen near real-time (Se#1). ML-services will therefore need to format the data on the go and provide smart services with the necessary input for event-based reasoning (Sy#3). While some smart services don't require event-based reasoning, they may however require a high level of security and dependability in their fulfillment (Se#4), while also satisfying high levels of compliance to certain other standards like Ethical, Legal and Social Implications (ELSI) (Sy#1). Smart Services such as the intelligent gatekeeper are necessary to provide access to the building for maintenance processes but also in emergency situations, they however also formulate requirements for high dependability, as is necessary to ensure access for authority's and emergency services. Hence, we categorize two types of services to address these different requirements (Se#3) for service providers to offer on the platform. These service types are demanding different sets of technology and processing for value creation. Therefore, in addition to supporting classical tightly coupled synchronous services (e.g. REST-service architectures) that require high reliability and fulfilment assurance, we suggest an asynchronous service infrastructure like a distributed event streaming component that can feed data into ML-models (Da#3) and deliver near real-time added value (Se#1).

4.2 Component description

In this section, we take a closer look at the individual components and their importance for a successful adaptation and operation of ML-based smart services in the ecosystem. In order to comply with the principle of an open ecosystem, the technical architecture must also follow an open design. This means that the components described below can be operated and resources integrated by different actors and that several solutions can be offered for the same function.

Scalable deployment infrastructure: Since real-time processing of large, high-frequency data volumes is required in the smart living data ecosystem (Se#1), the infrastructure should allow for flexible scaling. As such multiple actors should be able to add infrastructure resources to the data ecosystem. This feature is central to accommodate the pronounced heterogeneity of the data and the system landscape of smart living (Sy#3). The infrastructure needs to accommodate both asynchronous services, e.g. via message streaming, and synchronous services that (e.g. REST services) to be operated (Se#3). The use of

cloud resources can provide efficiency benefits, but compliance with data protection requirements to preserve the privacy of data owners must be considered (Da#1, Da#2, Da#4). Modern container application platforms offer the possibility to orchestrate and interlink services, so that instead of isolated monolithic services, recursive service structures are created on the platform (Se#2).

ML-training infrastructure: The operation of ML-based smart living services must necessarily include a component for training the various ML models (ML#1). This is a particular resource intensive taks as the models need to work on complex data sets, such as longer multivariate time series, images, or audio data. Various appropriate GPU cluster architectures exist [IMAK16] with potential different actors offering their infrastructure as workbenches in the platform ecosystem. The most appropriate choice for a specific service might depend on the specific use case and its associated data protection requirements. From a cost perspective, an virtual GPU cloud environment [HoSN17] might be efficient, but for many use cases processing personal data, such as movement patterns or health data, this might not be feasible (Da#1, Da#2, Da#4). The ecosystem component for ML training is therefore designed flexible and supports the integration of different environments by different actors, which can be selected in accordance with the use case (Sy#3).

Persistence: For true value-added solutions to be developed, ML-services must be built on multiple different data sources that were originally separate from each other. For example, an emergency detection service should be able to access occupant movement data, various sensor data (e.g., volume, temperature, humidity), and contextual information (e.g., time of day, weather data etc.). The underlying data types are obviously very heterogeneous, so there is a need for persistence that supports storage and linking across different sources and types (Da#6, Da#7). This persistence provides the data basis for the multivariate ML-services. Concrete components that should be included are, for example, time series databases for sensor data, graph databases for storing semantic relationships between objects in the home, key-value stores for image and sound data. To cope with the strong heterogeneity of the domain, a layer architecture that preprocesses and normalizes the raw data is reasonable. The goal must be to ensure semantic interoperability (Da#7). There are several efforts in the smart living domain to create uniform semantics, one of the more important of which is the W3C standard, which we propose to utilize. As the persistence is really hard to adjust for multiple actors to offer resources dynamically in the platform ecosystem we would propose a core group of actors in the ecosystem to reliably manage and build these components in a public and secure data environment such as the currently proposed GAIA-X [HKBB20].

Distributed event streaming component: To facilitate the interaction of services and event-based reasoning we propose to utilize a distributed event streaming component. The event streaming component will allow asynchronous ML-services to directly process the data from data owners like residents (Se#3). The component can process the large amounts of data that accumulate in the domain in a short time (Se#1) with integrity (Da#2). By preparing and providing higher level data and knowledge for other services they can work in conjunction extending the coverage of services for any given environment for real added value. By declaring their devices as things with the specific properties data owners can

connect and send their data into the event streaming component which will then be processed.

Registry/repository: Central registries and repositories create transparency across all actors about the services, models and data sources available in the data ecosystem (Sy#4). A concept of the service registry for smart living was introduced by Rebstadt et al. [RKHT21] and describes an architectural component that transparently provides meta-information about all services within a data ecosystem. A service registry offers service engineers with a tool that increases transparency within a data ecosystem and helps them develop privacy friendly (Da#1), explainable solutions, thereby contributing to domain requirements. To increase trust in the data ecosystem, the registry should, among other things, make transparent whether a service processes personal data, what risks might arise from its use and also support third-party certification [RKHT21]. In particular, when considering ML-services, stakeholders should be provided with information on the individual models and used data sources for training as well.

MLOps and continuous integration/Continuous development (CI/CD): MLOps describes a set of practices that aim at a reliable and efficient development, deployment, and maintenance of machine ML in production [ZhYD20]. These features contribute to the implementation of the domain requirement in terms of robustness (Sy#1), data integrity (Da#2) and traceability (Sy#2). The use of standardized pipelines with uniform data processing, test steps and logging of training results ensures that smart living ML-services have a certain quality before they are used in critical use cases by residents (ML#4). For this sub-area, explicit steps for uniform data (pre-)processing, data versioning, the actual training, the evaluation of the model and its deployment should be included. A logging of the training results and model parameters in a central model registry creates transparency and traceability (ML#5). It should be emphasized that, especially in the context of smart living, the definition of consumer-driven quality criteria and their consideration in the context of model evaluation are of crucial importance (Co#1). To this end, service consumers and their requirements for the quality of forecasts should be explicitly included in the development process. In the context of facial recognition, for example, it should be checked that the model used does not discriminate against certain subgroups of users based of their age, ethnicity, or skin color [KFRE22]. The CI/CD pipeline takes care of the deployment specifics, using templates for the different endpoints relevant in the domain. Automation and integrated testing reduce errors and increase the quality of the application and its robustness. After successful model deployment, the pipeline continuously collects data from its live operation and monitors performance. In this way, changes in user behavior are detected and can be compensated by retraining the model. Fig. outlines how a MLOps and CI/CD-pipeline helps integrate domain requirements into an automated and monitored development process.

Asynchronous ML-service architecture: The architecture of the ML-services is highly use case dependent but often they must be able to process the high frequency of the numerous sensor data in a performant way. At an IoT device level, the data owners (residents and the housing industry) may permanently supply personal data that provides information for instance about household activity, leakage, the ambient humidity and the health status

of relevant electrical appliances (e.g. dishwashers, washing machines etc.). For this purpose, an asynchronous ML-service architecture may be used, which consumes the service consumers' live data continuously for predictions. Their data sovereignty is protected, and they can authorize which personal data should be available for which use case and which provider. The raw data is then transferred to the Distributed Event Streaming Platform, where various transformation steps are performed, including a semantic annotation that ensures the standardization and interoperability of the raw data. The processed data is then consumed by various subsequent applications and stored in a persistence to build up an history for the purpose of training and logging. A potential consumer such as an authorized service provider could then use the data to train models for a service such as this asynchronous predictive maintenance service, as maintenance is a significant cost driver in the domain [LoMi21]. The trained model is then stored in the model registry, where it can be used again to build multiple services. A service using the model is then build and made available via the registry to potential users. For such asynchronous services based on the distributed event streaming platform, the data stream to be processed is first consumed via a consumer module. The module ensures that access is only granted to data that has been released to the application by the user. Sensor data needs to be deserialized and converted into an interpretable structure. Data attributes are determined via a filter and the preprocessing specific to the data (e.g. normalization or wavelet transformation) is performed, which processes the data for the ML-algorithm. Subsequently, the fitting pretrained ML-model based on the available data and attributes is selected and called in runtime, resulting in a prediction on the data or a virtual sensor output, combining the data from multiple sources into new knowledge. The results can then be processed and published. A standardized data structure (e.g. a JSON structure) should be selected, which simplifies further processing by subsequent services. Afterwards, the JSON object is serialized and transferred to the event streaming platform via a producer, where it can be consumed by subsequent services or added as a virtual sensor to an environment.

Authorization management (AM): To meet the domain requirements of ensuring data sovereignty (Da#4, Da#5) and data protection (Da#1), which is critical for data owners, the framework should offer a profound authorization management. [BHKH20], for example, have already presented a concept for the smart living domain. The AM needs to be consistent across all levels and components of the ecosystem (Da#2). The allocation of permissions must be granular (Da#5) and allow the data owner to grant selected actors in the ecosystem access to individual attributes of IoT devices or sensors and to revoke this access at any time (Da#4). To authorize the usage of data in the event streaming platform

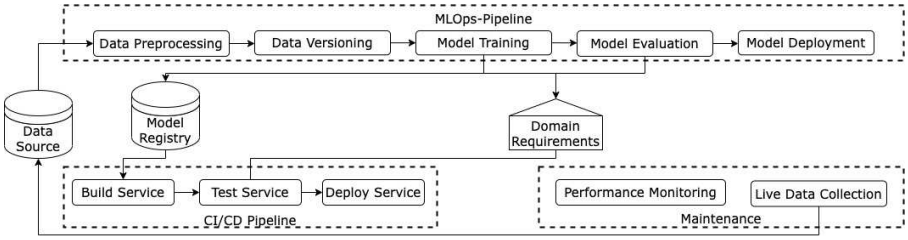


Fig. 4: CI/CD and MLOps on the platform

data owners themselves decide which data properties from which thing or sensor will flow into the platform in the IAM component (Da#5). For event-based reasoning all data from IoT devices are then provided in an interoperable IoT thing-schema (Da#7) on message topics. This however is not enough to secure the interest of data owners; the distributed event streaming component will need to ascertain that any data provided and distributed on different message topics can only be utilized by authorized services. For this a tight interaction between AM, service registry and event streaming component is necessary.

5 Discussion

The platform framework we propose is currently based on concrete requirements of the smart living data ecosystem, however, transferability offers great potential. It seems particularly suitable for ecosystems with many actors and a strong involvement of natural persons who contribute personal data as data owners, for example in the domains of health or mobility. Especially the fine-grained identity and access management, which provides for the assignment of authorization on the level of individual attributes in data streams, appears to be suitable for meeting demanding privacy requirements. From a value creation point of view, however, not only the extension of our framework to other ecosystems proves promising, but also a cross-ecosystem validation and adoption.

The value creation from data increasingly goes beyond internal boundaries of individual companies and disciplines and increasingly takes place in a cross-organizational and cross-domain environment [GMHF17]. In this way, potentials arising from asynchronous services, service registries and a high degree of scaling could be even better utilized and applied by a broader mass of actors, which could lead to an orchestration of completely new value propositions. In such a world, individual products increasingly take a back seat; instead, the value proposition shifts to data and its intelligent processing in smart service offerings [KaGK21]. Our ML oriented framework and especially the component of a granular identity and access management can help to protect this valuable resource: data and to strengthen the trust of the data owners in the ecosystem. The question arises as to who could operationalize such a framework. In its entirety, the implementation and operation of an architecture based on our platform framework exceeds the technical competencies and content scope of most of the actors in the smart living ecosystem. The housing industry is locally based and lacks the technical expertise to implement it. Product and service providers in the ecosystem tend to focus on their own product world. While large actors such as Hyperscalers like amazon or google might have the technical know-how, they do not offer the required level of trust or intent to enable a collaborative environment for the housing industry, energy providers, smart service providers and residents to develop an open ecosystem. The operation of the underlying platform infrastructure therefore appears to be an insoluble task for individual companies, which instead might only be overcome through cooperation between different actors with complementary competences. Such co-operation's also fits well into the two-sided market platform principle of our framework. It is worth noting that such considerations are stated goals of major international enterprise initiatives. European initiatives currently aim to create a federated data infrastructure for

Europe, to provide users with access to a broad specialized product portfolio and service catalogs, and thus promote the development of tailored solutions.

6 Conclusion and Future Work

To answer RQ1, in a first step, challenges of the smart living domain were identified from the literature as well as on the basis of real use case documents, leading to the derivation of requirements for the smart living ecosystem. Based on these domain requirements, this paper provides a platform framework for the development and operation of ML-services in the smart living data ecosystem and thus provides the answer to RQ2. Even though a broad spectrum of use cases and literature was included in the elicitation of the requirements for the framework, the practical evaluation is currently limited to individual use cases and components. A complete evaluation of the entire system and by a multitude of interdependent actors is still pending and should be the subject of future research in this area. One of the limitations of the proposed platform framework lies in its still very centralized structure as such it is currently not addressing the domain requirement Co#2 "support an open collaborative non centralized exchange of resources, data and services in the smart living ecosystem" in its full intended scope. This in itself shows a remarkable technical challenge and should be subject of its own future research. To provide the option for general execution of services on edge would require the integration of edge computing environments into the deployment scope, which is currently not covered in detail, and would also require further research as such the domain challenge ML#3 "On-edge executability of ML models" is not currently addressed either. Another open question for future research will be how to facilitate platform growth as it is bound to network externalities in the smart living ecosystem.

7 Acknowledgements

This research was conducted in the scope of the research project ForeSight, which is funded by the German Federal Ministry for Economic Affairs and Energy (BMWK). We would like to thank them for their support.

Bibliography

[BDMB17] Bianchini, D.; De Antonellis, V.; Melchiori, M.; Bellagente, P.; Rinaldi, S.: Data management challenges for smart living. In: Cloud Infrastructures, Services, and IoT Systems for Smart Cities: Springer, pp. 131–137, 2017.

[BHFR19] Bauer, J.; Hoffmann, H.; Feld, T.; Runge, M.; Hinz, O.; Mayr, A.; Förster, K.; Teske, F.; et al..: ForeSight - Platform Approach for Enabling AI-based Services for Smart Living. In: Pagán, J.; Mokhtari, M.; Aloulou, H.; Abdulrazak, B.; Cabrera, M. F. (eds.): How AI Impacts Urban Living and Public Health, Lecture Notes in Computer Science.

Bd. 11862. Cham: Springer International Publishing, pp. 204–211, 2019.

[BHKH20] Bauer, J.; Hechtel, M.; Konrad, C.; Holzwarth, M.; Hoffmann, H.; Feld, T.; Schneider, S.; Zinnikus, I.; et al.: ForeSight-An AI-driven Smart Living Platform, Approach to Add Access Control to openHAB. In: International Conference on Smart Homes and Health Telematics: Springer, pp. 432–440, 2020.

[Bowe09] Bowen, G.: Document Analysis as a Qualitative Research Method. Qualitative Research Journal, 9, pp. 27–40, 2009.

[BrHO06] Brewer, J.; Hunter, A.; et al.: Foundations of multimethod research: Synthesizing styles: Sage, 2006.

[BSSK17] Broring, A.; Schmid, S.; Schindhelm, C.; Khelil, A.; Kabisch, S.; Kramer, D.; Le Phuoc, D.; Mitic, J.; et al.: Enabling IoT Ecosystems through Platform Interoperability. In: IEEE Software 34/1, pp. 54–61, 2017.

[CBBF19] Chanson, M.; Bogner, A.; Bilgeri, D.; Fleisch, E.; Wortmann, F.: Blockchain for the IoT: Privacy-Preserving Protection of Sensor Data. In: Journal of the Association for Information Systems 20/9, 2019.

[ChHo19] Chang, F.; Ho, W.: An Analysis of Semi-Supervised Learning Approaches in Low-Rate Energy Disaggregation, 2019.

[GMHF17] Günther, W.A.; Mehrizi, M.; Huysman, M.; Feldberg, F.: Debating big data: A literature review on realizing value from big data. The Journal of Strategic Information Systems, Elsevier, pp. 191–209, 2017.

[GrHe13] Gregor, S.; Hevner, A.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly, 37, pp. 337–356, 2013.

[Grun20] Grundy, J.: Human-centric software engineering for next generation cloud-and edge-based smart living applications. In: 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing (CCGRID), IEEE, pp. 1–10, 2020.

[HiOS20] Hinz, O.; Otter, T.; Skiera, B.: Estimating Network Effects in Two-Sided Markets. Journal of Management Information Systems, 37, Routledge, pp. 12–38, 2020.

[HKBB20] Hoppe, G.; Kraemer, P.; Biegel, F.; Boll, S. (Hrsg.): GAIA-X A Pitch Towards Europe - Status Report on User Ecosystems and Requirements: Federal Ministry for Economic Affairs and Energy (BMWi), 2020.

[HoSN17] Hong, C.; Spence, I.; Nikolopoulos, D. S.: GPU virtualization and scheduling methods: A comprehensive survey. In: ACM Computing Surveys (CSUR) 50, ACM New York, NY, USA, 3, pp. 1–37, 2017.

[IMAK16] Iandola, F.N.; Moskewicz, M. W.; Ashraf, K.; Keutzer, K.: Firecaffe: near-linear acceleration of deep neural network training on compute clusters. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2592–2600, 2016.

[JKPG19] Jiang, J.; Kong, Q.; Plumbley, M.; Gilbert, N.: Deep Learning Based Energy Disaggregation and On/Off Detection of Household Appliances, 2019.

[JKPS19] Jung, D.; Kim, B.; Park, M.; Straub, D. W.: Innovation and Policy Support for Two-Sided Market Platforms: Can Government Policy Makers and Executives Optimize Both Societal Value and Profits? In: Information Systems Research, 30/3, INFORMS,

pp. 1037–1050, 2019.

[KaGK21] Kammler, F.; Gembarski, P.C.; Kortum, H.: Leveraging the Value of Data in the Continuum of Products and Services: Business Types in the Function-Oriented Offerings Model. In: Towards Sustainable Customization: Bridging Smart Products and Manufacturing Systems, Springer, pp. 773–780, 2021.

[KFRE22] Kortum, H.; Fukas, P.; Rebstadt, J.; Eleks, M.; Nobakht Galehpardsari, M.; Thomas, O.: Proposing a Roadmap for Designing Non-Discriminatory ML Services: Preliminary Results from a Design Science Research Project, 2022.

[KGZF20] Kortum, H.; Gravemeier, L.S.; Zarvic, N.; Feld, T.; Thomas, O.: Engineering of Data-Driven Service Systems for Smart Living: Application and Challenges. In: Lalic, B.; Majstorovic, V.; Marjanovic, U.; von Cieminski, G.; Romero, D. (eds.): Advances in Production Management Systems. Towards Smart and Digital Manufacturing, IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing, pp. 291–298, 2020.

[KoKM21] Kohl, T.A.; Kellner, D.; Mihale-Wilson, C. A.: Semi-supervised energy disaggregation for real-world adoption. In: Proceedings of the European Conference on Information Systems, virtual, 2021.

[KRHT22] Kortum, H.; Rebstadt, J.; Hagen, S.; Thomas, O. Integrating Data and Service Lifecycle for Smart Service Systems Engineering: Compilation of a Lifecycle Model for the Data Ecosystem of Smart Living. In: Proceedings of the 55th Hawaii International Conference on System Sciences, 2022.

[LoMi21] Lowin, M.; Mihale-Wilson, C.: Towards Predictive Maintenance as a Service in the Smart Housing Industry: Gesellschaft für Informatik, Bonn, 2021.

[MiRF19] Miori, V.; Russo, D.; Ferrucci, L. Interoperability of home automation systems as a critical challenge for IoT. In: 4th International Conference on Computing, Communications and Security (ICCCS). Rome, Italy: IEEE, 2019.

[NiDe15] Nikayin, F.; De Reuver, M.: What motivates small businesses for collective action in smart living industry? Journal of Small Business and Enterprise Development, Emerald Group Publishing Limited, 2015.

[OlLó18] Oliveira, M.; Lóscio, B.F.: What is a data ecosystem? In: Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, 2018.

[PTRC07] Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems, 3, pp. 45–77, 2007.

[RKGE22] Rebstadt, J.; Kortum, H.; Gravemeier, L.S.; Eberhardt, B.; Thomas, O.: Non-Discrimination-by-Design: Handlungsempfehlungen für die Entwicklung von vertrauenswürdigen KI-Services. HMD Praxis der Wirtschaftsinformatik, 59, pp. 495–511, 2022.

[RKHT21] Rebstadt, J.; Kortum, H.; Hagen, S.; Thomas, O.: Towards a transparency-oriented and integrating Service Registry for the Smart Living Ecosystem. In: INFORMATIK 2021, Berlin/Virtual, Germany. Gesellschaft für Informatik e.V., 2021.

[SCDM16] Sicari, S.; Cappiello, C.; De Pellegrini, F.; Miorandi, D.; Coen-Porisini, A. A security-

and quality-aware system architecture for Internet of Things. Information Systems Frontiers, 18, 4, pp. 665–677, 2016.

[SoGB13] Solaimani, S.; Guldemond, N.; Bouwman, H.: Dynamic stakeholder interaction analysis: Innovative smart living design cases. Electronic Markets, 23/4, pp. 317–328, 2013.

[SRLR19] Shin, C.; Rho, S.; Lee, H.; Rhee, W.: Data Requirements for Applying Machine Learning to Energy Disaggregation. Energies, 12, pp. 1696, 2019.

[VeTX16] Venkatesh, V.; Thong, J.; Xu, X.: Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. Journal of the Association for Information Systems, 17, pp. 328–376, 2016.

[WaGM17] Warkentin, M.; Goel, S.; Menard, P. Shared Benefits and Information Privacy: What Determines Smart Meter Technology Adoption? Journal of the Association for Information Systems, 18/11, 2017.

[WaYL18] Wang, P.; Yang, L.T.; Li, J.: An edge cloud-assisted CPSS framework for smart city. IEEE Cloud Computing, IEEE, 5, pp. 37–46, 2018.

[WuVS19] Wunderlich, P.; Veit, D.; Sarker, S.: Adoption of Sustainable Technologies: A Mixed-Methods Study of German Households. MIS Quarterly, 42, pp. 673–691, 2019.

[ZGIR12] Zoha, A.; Gluhak, A.; Imran, M.A.; Rajasegarar, S.: Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey. In: Sensors Bd. 12, Multidisciplinary Digital Publishing Institute, 12, pp. 16838–16866, 2012.

[ZhYD20] Zhou, Y.; Yu, Y.; Ding, B.: Towards MLOps: A Case Study of ML Pipeline Platform. In: International Conference on Artificial Intelligence and Computer Engineering (ICAICE). IEEE, pp. 494–500, 2020.

[ZNYL17] Zhang, K.; Ni, J.; Yang, K.; Liang, X.; Ren, J.; Shen, X.S.: Security and Privacy in Smart City Applications: Challenges and Solutions. IEEE Communications Magazine, 55/1, pp. 122–129, 2017.

[ZSBC21] Zhu, H.; Samtani, S.; Brown, R.; Chen, H.: A Deep Learning Approach for Recognizing Activity of Daily Living (ADL) for Senior Care: Exploiting Interaction Dependency and Temporal Patterns. MIS Quarterly, 45/2, pp. 859–896, 2021.

Designing a Data Sharing Tool Kit

Showing companies how to start using Data Sharing

Ilka Jussen [1], Julia Christina Schweihoff[2], Maleen Stachon[3] and Frederik Möller[1,4]

Abstract: Sharing data is essential to the success of modern data-driven business models. They play a crucial role for companies in creating new and better services and optimizing existing processes. While the interest in data sharing is growing, companies face an array of challenges preventing them from fully exploiting data sharing opportunities. Mitigating these risks and weighing them against their potential is a creative, interdisciplinary task in each company. The paper starts precisely at this point and proposes a Tool Kit with three Visual Inquiry Tool (VIT) to work on finding data sharing potential conjointly. We do this using a design-oriented research approach and contribute to research and practice by providing three VITs that help different stakeholders or companies in an ecosystem to visualize and design their data-sharing activities.

Keywords: Data Sharing, Visual Inquiry Tools, Canvas Development, Design Science Research.

1 Introduction

The handling and use of data is a controversial topic. On the hand, it seems clear that sharing data has a high potential to improve data-based ecosystems [Az20]. On the other hand, companies face various challenges (e.g., trust issues) that make them reluctant to share data [Ge21, Op21]. Subsequently, the use of data as a core resource for designing data-driven business models [Ha14] is getting more critical by the day. Nowadays, data can help to enable more sustainable supply chains by sharing it between actors to reduce the carbon footprint jointly [Ga22].

A current study by the European Commission expects constant growth in the volume of data in Europe. The volume is expected to increase from 33 zettabytes in 2018 to 175 zettabytes in 2025. In addition, 80% of industrially generated data is not used, showing the potential already available [Eu22]. The European Commission estimates the value of

[1] Lehrstuhl für industrielles Informationsmanagement/ Technische Universität Dortmund, Joseph-von-Fraunhofer-Straße 2-4, 44227 Dortmund, ilka.jussen@tu-dortmund.de, https://orcid.org/0000-0002-0498-2043
[2] Lehrstuhl für industrielles Informationsmanagement/ Technische Universität Dortmund, Joseph-von-Fraunhofer-Straße 2-4, 44227 Dortmund, julia.schweihoff@tu-dortmund.de
[3] Lehrstuhl für industrielles Informationsmanagement/ Technische Universität Dortmund, Joseph-von-Fraunhofer-Straße 2-4, 44227 Dortmund, maleen.stachon@tu-dortmund.de
[4] Fraunhofer Institute für Software- und Systemtechnik, Joseph-von-Fraunhofer-Straße 2-4, 44227 Dortmund, Frederik.Moeller@isst.fraunhofer.de

doi:10.18420/inf2022_33

the data economy at the end of 2019 at around 324.86 billion euros. This corresponds to approximately 2.6% of the gross domestic product (GDP) of the 27 member states of the European Union (EU-27) [Mi21]. Here, it is forecasted that the value of the data economy will grow to an estimated 829 billion euros within the EU-27 states by 2025 [Eu21].

At the same time, the monetization of data and the development of data-driven business models pose great challenges for many companies due to the unclear financial measurement of the value of their data [Pa21]. Other factors also influence the willingness of many companies to share data, as a study by the Institute of the German Economy from 2020 shows. This study shows the variety of obstacles that currently exist. These include concerns about unauthorized data use and ambiguities in data protection law [In21]. The need to create a secure space for data sharing is also reflected in creating the European initiative Gaia-X. The initiative by European countries pursues the goal of creating a secure and trustworthy framework for data sharing to ensure sustainable advantage between several involved actors [GA21].

Against this background, researchers and practitioners from different companies in an ecosystem require tools to assist them in uncovering the potential of data sharing. In the paper, we propose a tool kit using Visual Inquiry Tools (VIT) or Canvases (e.g., see the Business Model Canvas [OP10]) to assist in solving these issues. Canvases are excellent tools to work creatively and conjointly in interdisciplinary teams on a given problem area ([EP19, TMB19]). Based on this, the need to develop a thematically appropriate canvas is therefore purposeful. With the help of the developed canvases, we want to answer the following research question: *How to bring together interdisciplinary stakeholders in data sharing projects?*

2 Theoretical Background

2.1 Data sharing

There is no comprehensive definition of *'data sharing'* in the literature. Instead, the term is often used synonymously with *'data exchange'*. Table 1 shows some of the approaches to defining data sharing. According to [NST19], data sharing involves exchanging data between two or more parties. [VHT19] also focus on data access in data sharing and emphasize the possibility of using data to develop innovations and services. An additional focus in the conceptual definition is given by [Ma20], who refers to the different financial modalities of data sharing. Data can be shared either free of charge, for a financial fee, or for other data. In addition, they distinguish between direct data sharing, in which data is shared one-to-one, and indirect data sharing in the form of data-based services.

Definitions	Sources
"In this article, we define data sharing as an exchange of data between different stakeholders, and when happening on a platform, as lateral data exchange in a network."	[NST19]
"Data sharing can allow organizations to access complementary data sources and help them develop innovative applications and services. We considered data sharing as either third parties' opening the data they own or their consuming the data from other providers."	[VHT19]
" Data sharing is a label that may cover different economic modalities: sharing for free, trading for a monetary compensation or in exchange for other data, direct sharing of a dataset or indirect sharing of a data-based service only."	[Ma20]

Tab. 1 Definitions of data sharing

Companies can own the data and, at the same time, share it with others to jointly solve business problems that the companies could not solve on their own. In this case, the companies learn to extract previously unknown information from the data [TWS19, vv15]. Companies can also learn to understand their internal processes better from data [Ga21]. Ideally, data sharing can lead to the development of creative ideas or new products, both for the data owner and the data user [TWS19]. One example is that monetizing data via data marketplaces or in direct cooperation with other companies can reveal a lucrative opportunity for companies [Ga21]. Another benefit is the shared increase in data quality, making it easier for the data user to use the data and improve it for the data owner [TWS19, DN19]. In general, the intensive use and sharing of data can lead companies, both data owners and users, to increase their capabilities regarding the use of data. This can happen, for example, through new capabilities linked to machine learning or artificial intelligence [TWS19]. To exploit the potential of data sharing, companies need to participate in an ecosystem. The incentives must be clearly communicated to companies to alleviate their previous concerns [Ge21].

In general, it becomes evident from our literature analysis that the current image of data sharing is associated more with the potential risks than with its opportunities. The fear of disclosing too much sensitive data plays just as big a role as the concern about hacking attacks [DN19, GO20]. In this context, it becomes clear how significant the worries are about losing control over one's data or not being able to prevent possible data misuse on the part of users [Ba19, vv15]. The fear of losing intellectual property and knowledge to other companies leads to a lack of openness on the part of companies [GO20]. Among others, [ADR21] highlight the competitive advantages for other companies or competitors and fears of disclosing sensitive end-user data [Ga21, Ge21]. All the risks above can be justified by a lack of trust between the companies that own data and other actors within the ecosystem [DN19, KKT14]. This mistrust is evident between companies that already had business relationships with each other before the sharing of data [GO20]. Another risk for companies is the unclear costs that can arise during preparations and the

implementation of the data sharing [DN19]. The unclear financial valuation of their own data also poses major challenges for companies [MSO20, Pa20].

2.2 Visual inquiry tool design

VITs provide a collaborative space (often two-dimensional) to generate a 'common' playing field for teams to work on solving problems, developing new ideas, or designing new solutions [Av20, EP19]. Other terms commonly used to describe VITs are 'design canvas' (e.g., [TMB19]) or 'visual collaboration tool' (e.g., [Sa00]). The paper develops multiple VITs for inter-organizational data sharing.

To do this, we rely on two foundations. First, we use the design principles for VITs by [Av20] and, second, the design taxonomy for VITs by [MS22]. The iterative and interdisciplinary design process is ideally suited to the topic of data sharing. Opinions of different people who take on different roles in the company must be considered. A canvas lends itself to presenting complex and sometimes unclear topics or issues in a more understandable way. Furthermore, these tools are suitable for use both in academia and in practice as well as in the business environment [Av20, CN18]. Additionally, the collaborative process of idea development, both verbal and non-verbal communication in interdisciplinary groups, is a reason for its versatile applicability [Av20, MSO21]. For the development of the canvas, we followed the design principles of [Av20]. Twelve design principles imply the three main principles of 'conceptual model', 'shared visualization' and 'instructions for use'. In addition to the design principles of [Av20], the taxonomy of [MS22] is used in this work on the topic of the design options of VITs. This taxonomy offers helpful characteristics regarding the aspects of the problem definition, the design process, the solution, and the final evaluation of the developed tools.

3 Research Design

Given that VITs are typically a product of design-oriented research (see [MS22, Sc21]), we use the *'Design Science Research Methodology Process Model'* [Pe07] (see Fig. 1). The method is suitable for developing and evaluating innovative artifacts in information technology (IT) and solving organizational problems [He04]. The method of [Pe07] consists of six steps, of which the first five will be explained in more detail, as the sixth step, *'Communication'* is the publication of this paper itself. Overall, the method is iterative. Accordingly, the steps are critically evaluated and questioned repeatedly to constantly improve the content and the knowledge gained in this work.

The first step is to *'Identify Problem and Motivation'*. The potential of data sharing as a part of data-driven business models identified in the literature research is the motivation for our paper (see Sections 1 and 2). On the other hand, the lack of concepts that focus on the positive aspects of data sharing is identified as a problem.

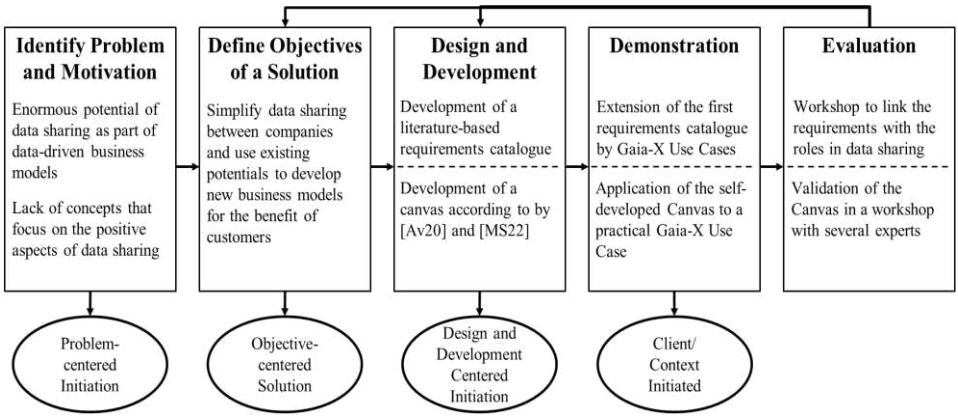


Fig. 1: Design Science Research Methodology according to [Pe07]

The second step of the method focuses on *'Define Objectives of a Solution'* for the intended solution. The research aims to simplify data sharing between companies by developing a helpful artifact and enabling new cooperation between companies or joint business models. These should use data sharing possibilities to create greater added value for customers than was previously the case. Depending on what insights are gained in the following steps, this step is iteratively conducted. The objectives are constantly redefined and more specific to focus on the work as precisely as possible. Following standard practice in DSR, we first develop a requirements catalogue for data sharing activities based on a literature review. Please note that the requirements catalogue is not explained in detail due to spacing limitations. In step 3 *'Design and Development'*, we propose three VITs, that address these requirements based on the roles they are associated with (e.g., data provider, data consumer, and orchestrator). For this purpose, the findings of [Av20] and [TMB19, TMB20] are used as guidance from the literature (see Section 2). This step aims to develop the first version of a canvas that will be iteratively refined in the future. In the fourth step, *'Demonstration'*, the canvas is applied to a practical use case to see, based on the practical application, areas where there is still a need for change. In the last step, *'Evaluation'*, the canvas is applied in a workshop with experts from various field. Here, the participants' external perspective should help evaluate and identify further suggestions for changes to the existing canvas. Through the iterative procedure, the individual interim results are critically analyzed as much as possible and constantly evaluated to bundle the collected findings in a canvas.

4 A Tool Kit for Data Sharing

4.1 Developing and evaluating the tool kit

To identify relevant requirements for data sharing between different actors, we conducted a literature review according to [WW02]. For this purpose, we searched for the terms *'data sharing'* and *'requirements'* in the title, abstract, and keywords in various combinations on the AiSeL and Scopus databases. In this literature review, we identified 24 papers as helpful, which state requirements that need to be considered for the different roles in data sharing. The relevant requirements that need to be defined at the beginning of data sharing between different actors include aspects such as data sovereignty, data security, and data quality. The first step was to develop a canvas tool kit with relevant building blocks and questions for the three essential roles of a data ecosystem, according to Oliveira and Lóscio (***data provider, data consumer, and intermediary***) [OL18], which were also classified as relevant for data sharing. This canvas consisted of 10 building blocks that included topics such as data sovereignty, data quality, and data value. These ten building blocks were assigned to the superordinate areas of value proposition and data sharing environment. Pictograms and arrows were used in addition to the headings and questions of the building blocks to make it easier for the user to use the canvas. This also represents a design principle according to Avdiji et al. by thematizing the common visualization [Av20].

The first Canvas was tested with academic staff researching in the field of data sharing. For this purpose, a partial scenario for a Gaia-X [5] use case was developed as an example to test the application of the Canvas using a real-life example. The testers evaluated the use of the Canvas using the Likert scale to provide feedback to the researcher on how the Canvas can be further developed [Li32]. Among the key findings of this evaluation is that it was recommended to develop three separate canvases for the respective roles. Furthermore, the individual building blocks should be questioned to see if all building blocks are equally relevant for the three roles. In addition, the corresponding questions on the building blocks should be adapted to the perspective of the respective role. After this first round of evaluation, the canvas was fundamentally changed based on the feedback from the users. In this way, the three canvases as a tool kit were created that were classified as important for the three roles in data sharing. The number of building blocks was also reduced. Therefore, the consumer canvas has six building blocks, the producer canvas has five building blocks, and the mediator canvas has four building blocks.

[5] Gaia-X Use Case 'Collaborative Condition Monitoring': https://www.bmwi.de/Redaktion/EN/Artikel/Digital-World/GAIA-X-Use-Cases/collaborative-condition-monitoring.html

4.2 The data sharing tool kit

Below, we explain the three canvases in more detail (see Fig. 2, 3, and 4) for the respective roles. The canvases are shown in a filled-out version with the help of the Gaia-X *(Collaborative Condition Monitoring)* use case based on the evaluation. We have developed three canvases, as it became apparent during the development process that the stakeholders consider similar but not the same issues to be important. The consumer's canvas consists of the six building blocks: *Data Interest, Purpose of Data Use, Data Sovereignty, Expected Data Quality, Data Preservation,* and *Data Value*. The first building block, ***'Data Interest',*** contains which data the consumer would like to have from the provider. This can be data the provider is already collecting or data that the provider could collect with additional effort. The building block ***'Purpose of Data Use'*** is only classified as necessary for the role of the consumer. The consumer should indicate to the provider the purpose for which the data provided is to be used. The provider, in turn, only must agree to this purpose. Accordingly, this would only be a yes/no decision by the producer depending on the intended use on the part of the consumer. In addition, the building block of ***'Data Sovereignty'*** was found to be relevant to the present context.

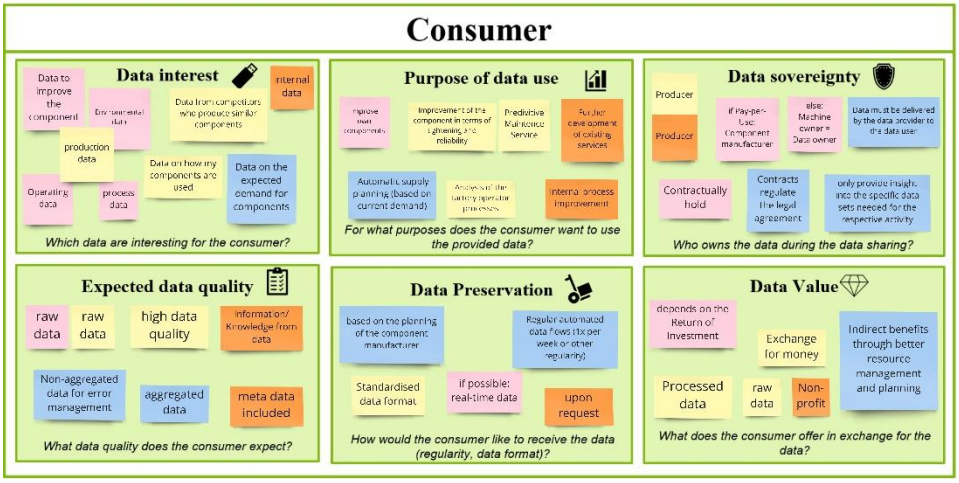


Fig. 2: Data Sharing Canvas - role: consumer (use-case based evaluation)

This implies a highly relevant question for data sharing, i.e., which actor is the owner of the data while the consumer is using it. Maintaining sovereignty over one's own data can counteract the current risks that companies see in sharing data. Furthermore, ***'Expected Data Quality'*** defines another building block of the developed canvas. Certain parameters must be defined for the data to be useful to the final consumer. These include aspects such as the inclusion of metadata or the integrity of the data, which implies the correctness and completeness of the data. Besides the building blocks already mentioned, the building block of ***'Data Preservation'*** should be jointly defined within the data-sharing framework. This component includes aspects such as the temporal continuity of the provision and the

data format in which the data is shared. Factors such as the provision of data in data catalogs and the portability of these should also be considered in this context. The value proposition's sixth and final building block is ***'Data Value'***. As already mentioned above, this represents a major challenge, or rather even a risk. The danger of passing on data to others for too low a price characterizes the image of companies.

The producer's canvas in Fig. 3 consists of the fifth building blocks: *Data Availability, Existing Data Form, Data Provision, Data Sovereignty, and Data Value*. The building block ***'Data Availability'*** implies which data the provider is willing to share with the consumer. Here, it is expected that there is sensitive internal company data that stakeholders do not want to share, unlike specific process data. Furthermore, the building block ***'Existing Data Form'*** addresses the quality of the data available to the producer. The quality of the data can have a massive influence on the usability of the consumer. Subsequently, the building block ***'Data Provision'*** implies how the data is provided to the consumer. Questions such as the frequency of data provision or the data format used should be clarified here. As for the consumer, the ***'Data Sovereignty'*** building block is also relevant for the producer and the intermediary. All three actors must also come to a common understanding of the data's owner during data sharing. The situation is like the building block ***'Data Value'***. Here, what the consumer offers for the data must match what the producer expects in terms of compensation for providing it.

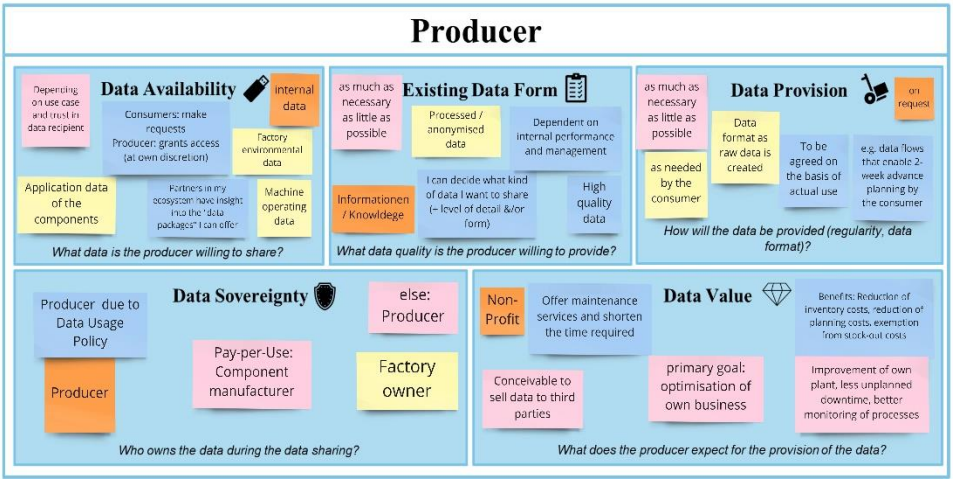


Fig. 3: Data Sharing Canvas - role: producer (use-case based evaluation)

The intermediary's canvas in Fig. 4 consists of the four building blocks: *Data Sovereignty, Remuneration, Platform, and Data Security*. As already mentioned, ***'Data Sovereignty'*** is also important for the intermediary. The building block ***'Remuneration'*** means that the intermediary must state how he/she wishes to be remunerated for his/her services. This can be done, for example, by a percentage of financial participation for each data sharing process. Another building block is ***'Platform'.*** The intermediary would adapt this to the

requirements of consumers and providers. The last building block is ***'Data Security',*** by which we mean how the intermediary wants to secure the data sharing process in the sense of the actors involved. In this context, questions regarding securing data access for authorized persons only should be considered.

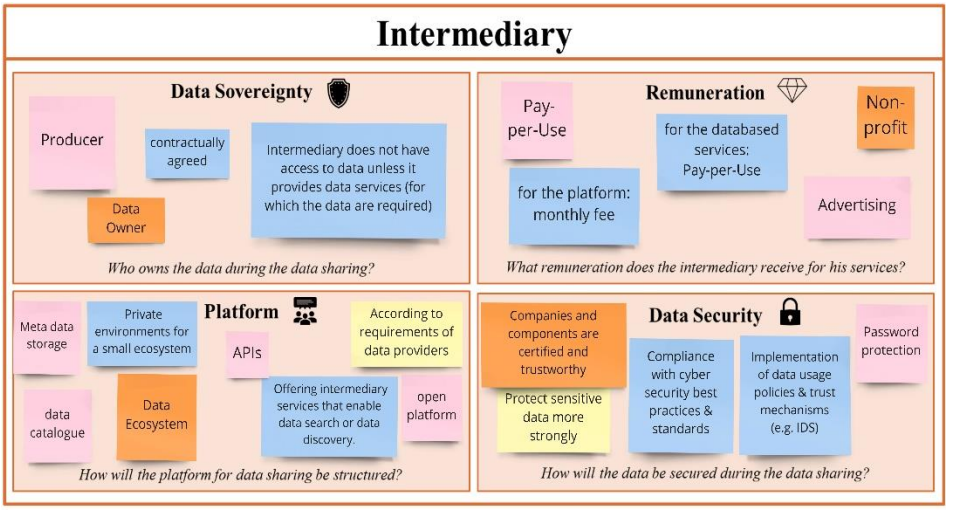


Fig. 4: Data Sharing Canvas - role: intermediary (use-case based evaluation)

The tool kit aims to facilitate collaboration between interdisciplinary groups on data sharing and wants to ensure that the different roles clearly define their position before the start of the inter-organizational data sharing. Users should ask themselves the following questions: *What is important to me in sharing my data with other actors?* Depending on their role in the data sharing process, they should fill in the respective canvas. The components of the canvases reflect some aspects that are important for data sharing across all sectors, regardless of the specific use case. If actors take on more than one role, they should also fill more than one canvas. Thus, with the help of the tool kit, a common understanding is to be established before the data sharing begins to prevent misunderstandings that could negatively influence trust among each other.

5 Contributions, Limitations, and Outlook

Even though the potential of data sharing is sufficiently well known, it often still fails to be implemented because there is often neither trust nor transparency between the companies. To prevent this in the future, tools are needed to support this sharing between the different companies, some of which we **contribute** here. Inter-organizational data sharing allows companies to use data as a resource to drive the new and further development of their own processes and services [RS19]. A variety of potentials through

data sharing have been identified in this work. These include the development of new services and products or learning from others. On the other hand, concerns and risks were repeatedly associated with data sharing. Aspects such as the loss of control or doubts about the security of one's own data influence the openness of companies towards inter-organizational data sharing. In the course of this work, it became clear that data sharing is not to be considered in isolation but is strongly linked to other large blocks of topics, such as data law, data value, and data sovereignty. The artifact users are companies interested in data sharing, which so far lack a guideline that implies the relevant aspects that need to be defined before data sharing begins.

To be able to classify the results in a differentiated manner regarding their significance, we now discuss its **limitations**. One limitation of this work is the perspective chosen for its implementation. The topic of inter-organizational data sharing was viewed from an economic and business model-focused perspective. The information technology perspective, which is interesting in this context, was neglected in this work and should be included in future research. The tool kit was developed based on the researcher's opinion. Some requirements were found to be more significant and therefore form their own building blocks. The researcher gave the decision on which building blocks or underlying requirements are relevant in the first step and then influenced by the testers. Furthermore, a differently chosen use case could possibly focus on other building blocks. In addition, it should be noted that the developed tool kit can serve as a tuning tool before the start of the data sharing and accordingly cover a subprocess rather than the entire data sharing.

After the intensive examination of inter-organizational data sharing, **further research** will now be considered that was not taken up in this work due to the limited scope. The evaluation in a setting with companies representing each role would enhance the validity of the canvases. What could emerge from further developing and evaluating the canvases is a need for more role-specific canvases (e.g., in scenarios where there are more detailed roles or specific types of roles or data to be shared). These can also be sector-specific roles that are necessary for the specific application. But not only the number of canvases can be expanded in the future, but also the variety of canvases about their time of application. In this work, three canvases were developed that can be used as a voting medium before data sharing begins. Accordingly, the development of further tools suggests itself, which could, for example, be used again and again during the data sharing as a basis for the further planning of the inter-organizational data sharing. Questions can be focused on how existing activities for data sharing between different actors can be further intensified.

6 Acknowledgements

This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the Incentives and Economics of Data Sharing Funding Action (IEDS0001). The author is responsible for the content of this publication.

Bibliography

[ADR21] Agahari, W.; Dolci, R.; Reuver, M. de: Business Model Implications of privacy-preserving technologies in data marketplaces: The case of multi-party computation. In: Proceedings of the 29th European Conference on Information Systems, 2021.

[Av20] Avdiji, H. et al.: A Design Theory for Visual Inquiry Tools. Journal of the Association for Information Systems 3/21, pp. 1–40, 2020.

[Az20] Azkan, C. et al.: SERVICE DOMINANT LOGIC PERSPECTIVE ON DATA ECOSYSTEMS - A CASE STUDY BASED MORPHOLOGY. In: Proceedings of the 28th European Conference on Information Systems, 2020.

[Ba19] Bastiaansen, H. et al.: Infrastructural Sovereignty over Agreement and Transaction Data ('Metadata') in an Open Network-Model for Multilateral Sharing of Sensitive Data. In: Proceedings of the International Conference on Information Systems, 2019.

[CN18] Chandra Kruse, L.; Nickerson, J.: Portraying Design Essence. In: Proceedings of the 51st Hawaii International Conference on System Sciences, pp. 4433–4442, 2018.

[DN19] Dahlberg, T.; Nokkala, T.: Willingness to Share Supply Chain Data in an Ecosystem Governed Platform – An Interview Study. In: BLED Proceedings, pp. 619–638, 2019.

[EP19] Elikan, D.; Pigneur, Y.: A Visual Inquiry Tool for Brand Identity. In: Proceedings of the 52nd Hawaii International Conference on System Sciences, pp. 571–580, 2019.

[Eu21] Europäische Kommission: Europäische Datenstrategie. Die EU zum Vorbild für eine digitale Gesellschaft machen. https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_de, Stand: 28.08.2021.

[Eu22] European Commission: Data Act: Commission proposes measures for a fair and innovative data economy. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113, Stand: 16.03.2022.

[Ga21] Gabriel, P.: How to share data? Data-Sharing-Plattformen für Unternehmen. Betriebswirtschaftliche und juristische Grundlagen, aktuelle Praxisprojekt, erste Handlungsempfehlungen. https://www.datasharing-sdw.de/story/SDW_Studie_DataSharing_DE.pdf, Stand: 12.11.2021.

[GA21] GAIA-X: What is Gaia-X. https://www.gaia-x.eu/what-is-gaia-x, Stand: 01.09.2021.

[Ga22] Galer, S.: For High Tech Supply Chains, CO2 Emission Tracking Is Sustainable Business Currency. https://www.forbes.com/sites/sap/2022/04/19/for-high-tech-supply-chains-co2-emission-tracking-is-sustainable-business-currency/?sh=3fcb329f356f, Stand: 27.04.2022.

[Ge21] Gelhaar, J. et al.: Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. In: Proceedings of the Pacific Asia Conference on Information Systems, pp. 1–14, 2021.

[GO20] Gelhaar, J.; Otto, B.: Challenges in the Emergence of Data Ecosystems. In: Proceedings of the Pacific Asia Conference on Information Systems, 2020.

[Ha14] Hartmann, P. M. et al.: Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms, 2014.

[He04] Hevner, A. R. et al.: Design Science in Information Systems Research. MIS Quarterly 1/28, pp. 75–105, 2004.

[In21] Institut der deutschen Wirtschaft Köln: Datenwirtschaft in Deutschland. Wo stehen die Unternehmen in der Datennutzung und was sind ihre größten Hemmnisse? https://bdi.eu/publikation/news/datenwirtschaft-in-deutschland/, Stand: 06.01.2022.

[KKT14] Knol, A.; Klievink, B.; Tan, Y.-H.: Data Sharing Issues and Potential Solutions for Adoption of Information Infrastructures: Evidence from a Data Pipeline Project in the Global Supply Chain over Sea. In: BLED Proceedings, pp. 1–12, 2014.

[Li32] Likert, R.: A technique for the measurement of attitudes. Archives of Psychology 140/22, pp. 1–55, 1932.

[Ma20] Martens, B. et al.: Business-to-Business data sharing: An economic and legal analysis. https://ec.europa.eu/jrc/sites/default/files/jrc121336.pdf, Stand: 17.09.2021.

[Mi21] Mildebrath, H.: Data Governance Act. Briefing - EU Legislation in Progress. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690674/EPRS_BRI(2021)690674_EN.pdf, Stand: 21.08.2021.

[MS22] Möller, F.; Steffen, B.: 'License to VIT' - A Design Taxonomy for Visual Inquiry Tools. In: Proceedings of the Hawaii International Conference on Systems Sciences, 2022.

[MSO20] Mosch, P.; Schweikl, S.; Obermaier, R.: Trapped in the supply chain: The digital Servitization of an industrial Supplier - out of the frying pan into the fire? In: Proceedings of the 28th European Conference on Information Systems, 2020.

[MSO21] Möller, F.; Schoormann, T.; Otto, B.: ‘Caution – Principle Under Construction’ A Visual Inquiry Tool for Developing Design Principles. In: 16th International Conference on Design Science Research in Information Systems and Technology (DESRIST), Kristiansand, Norway, 2021.

[NST19] Nokkala, T.; Salmela, H.; Toivonen, J.: Data Governance in Digital Platforms. In: Proceedings of the Americas Conference on Information Systems, 2019.

[OL18] Oliveira, M. I. S.; Lóscio, B. F.: What is a data ecosystem? dg.o 2018. In: Proceedings of the 19th Annual International Conference on Digital Government Research, Delft, Netherlands. ACM, New York, NY, USA, pp. 1–9, 2018.

[OP10] Osterwalder, A.; Pigneur, Y.: Business Model Generation: Ein Handbuch für Visionäre, Spielveränderer und Herausforderer. Campus Verlag, Frankfurt, New York, 2010.

[Op21] Opriel, S. et al.: Requirements for Usage Control based Exchange of Sensitive Data in Automotive Supply Chains. In (Bui, T. ed.): Proceedings of the 54th Hawaii International Conference on System Sciences, 2021.

[Pa20] Parvinen, P. et al.: Advancing Data Monetization and the Creation of Data-based Business Models. Communications of the Association for Information Systems 47, pp. 25–49, 2020.

[Pa21] Panett, K.: 3 Ways to Monetize Data and Analytics. https://www.gartner.com/smarterwithgartner/3-ways-to-monetize-data-and-analytics/, Stand: 20.08.2021.

[Pe07] Peffers, K. et al.: A Design Science Research Methodology for Information Systems Research. Journal of Management Information Systems 3/24, pp. 45–77, 2007.

[RS19] Richter, H.; Slowinski, P. R.: The Data Sharing Economy: On the Emergence of New Intermediaries. IIC - International Review of Intellectual Property and Competition Law 50, pp. 4–29, 2019.

[Sa00] Sanders, E.-N.: Generative Tools for CoDesigning. Collaborative Design, 2000.

[Sc21] Schoormann, T. et al.: On Your Mark, Ready, Search: A Framework for Structuring Literature Search Strategies in Information Systems. In: Proceedings of the 16th International Conference on Wirtschaftsinformatik, pp. 1–15, 2021.

[TMB19] Thoring, K.; Müller, R. M.; Badke-Schaub, P.: Exploring the Design Space of Innovation Canvases. In: Conference Proceedings of the Academy for Design Innovation Management, pp. 1–13, 2019.

[TMB20] Thoring, K.; Mueller, R.; Badke-Schaub, P.: Assessing and Changing an Organization's Innovation Culture with the Workspace Catalyst Canvas. In: Proceedings of the Hawaii International Conference on System Sciences, 2020.

[TWS19] Thuermer, G.; Walker, J.; Simperl, E.: Data sharing toolkit: Lessons learned, resources and recommendations for sharing data. https://datapitch.eu/wp-content/uploads/2019/10/7770-Final-Data-Sharing-Toolkit-Web.pdf, Stand: 17.09.2021.

[VHT19] Vesselkov, A.; Hämmäinen, H.; Töyli, J.: Design and Governance of mHealth Data Sharing. Communications of the Association for Information Systems 18/45, pp. 299–321, 2019.

[vv15] van den Broek, T.; van Veenstra, A. F.: Modes of Governance in Inter-Organizational Data Collaborations. In: Proceedings of the 23rd European Conference on Information Systems, 2015.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 2/26, pp. xiii–xxiii, 2002.

Using Machine Learning to Predict POI Occupancy to Reduce Overcrowding

Jessica Bollenbach [1], Stefan Neubig [2], Andreas Hein[3], Robert Keller [4] and Helmut Krcmar[5]

Abstract: Due to the rapid growth of the tourism industry, associated effects like overcrowding, overtourism, and increasing greenhouse gas emissions lead to unsustainable development. A prerequisite for avoiding those adverse effects is the prediction of occupancy. The present study elaborates on the applicability and performance of various prediction models by taking a case study of beach occupancy data in Scharbeutz, Germany. The case study compares different machine learning models once as supervised machine learning models and once as time series models with a persistence model. XGBoost and Random Forest as time series demonstrate the most accurate prediction, followed by the supervised XGBoost model. However, the short prediction span of time series models is a disadvantage for longer-term visitor management to avoid the explained unsustainable effects through steering measures, so depending on the use case, the XGBoost model is to be favoured.

Keywords: Beach Occupancy, Time series Forecast, XGBoost, Random Forest, Support Vector Regression, SARIMA, Tourism Demand.

1 Introduction

In recent decades, the international tourism sector has experienced significant growth from 25 million cross-border arrivals in 1950 to more than 1.3 billion arrivals in 2017, which is expected to continue up to 1.8 billion tourists in 2030 [Wo18]. Despite all the benefits, this upward trend is also driving unsustainable tourism, with effects such as overtourism, local overcrowding, and increasing greenhouse gas emissions due to the close link of tourism and mobility [Ca19], [Hø00]. Most leisure trips are made by greenhouse gas emitting cars, as touristic points of interests (POIs) in rural areas are often poorly accessible with public transport and people seek self-determination in their leisure time [Ei22], [GG18]. Overcrowding effects at the POI increase these emissions due to the

[1] University of Applied Sciences Kempten, WTZ, Bahnhofstraße 61, 87435 Kempten (Allgäu), & Fraunhofer Institute for Applied Information Technology FIT, Augsburg, jessica.bollenbach@hs-kempten.de, https://orcid.org/0000-0001-9554-1640
[2] Technical University of Munich, Boltzmannstr. 3, 85478 Garching, & Outdooractive AG, Missener Str. 18, 87509 Immenstadt, stefan.neubig@tum.com, https://orcid.org/0000-0002-3794-7260,
[3] Technical University of Munich, Boltzmannstr. 3, 85478 Garching, andreas.hein@tum.de
[4] University of Applied Sciences Kempten, WTZ, Bahnhofstraße 61, 87435 Kempten (Allgäu), & Fraunhofer Institute for Applied Information Technology FIT, Augsburg, robert.keller@hs-kempten.de, https://orcid.org/0000-0001-7097-1724
[5] Technical University of Munich, Boltzmannstr. 3, 85478 Garching, helmut.krcmar@tum.de

doi:10.18420/inf2022_34

increased chance of congestion and long search time for a parking lot [Pa22]. In addition, a lack of parking lots at overcrowded POIs lead to environmentally harmful illegal parking. Further, the mass of people at overcrowded POIs are responsible for pollution, noise, and wildlife disturbance. Such overcrowding effects occur especially in free public spaces like a beach, lake, city center, or mountain trail, as access cannot be controlled since no closure during peak occupancy occurs. Further, no final and predefined maximum occupancy exists, unlike, for example, in a parking garage. This leads to uncontrolled peak occupancy which is unsustainable and unpleasant for any visitor. To establish a way to more sustainable tourism, Schmücker et al. [Sc22] propose the demand for an active visitor management (AVM) system which guides visitors to more sustainable behavior. For example, steering measures enabled by an AVM system target the even distribution of visitors between open accessible POIs to reduce overcrowding or target the controlled filling of parking lots to reduce congestion and search time. A key element of such an AVM system is the prediction of POI occupancy because only with this knowledge can effective and target-oriented steering measures be initiated. POI occupancy prediction, or more generally, tourism demand prediction, is a well-established research field [WSS17]. However, existing and utilized data often lacks information on day tourism and tends to focus on overnight stays, thus creating an incomplete picture [Ne22]. Furthermore, the considered geographical areas are mainly larger regions [YZ19] or smaller, mostly closed areas like parking lots [APB21] and lack the examination of (semi-) open terrain of a specific POI. Hence, occupancy prediction models need to be developed for several days in advance to enable steering measures in an AVM system. These models need to cover a fine-grained temporal observation of a (semi-) open accessible terrain, including day visitors and overnight stays, which is, to the best of our knowledge, currently a research gap.

In this work, we evaluate different prediction model types, including various machine learning (ML) models, to predict POI occupancy of a local, semi-open terrain. The development of the prediction model is based on a use case study with occupancy data collected via sensors of the Bay of Lübeck in Scharbeutz at the Baltic Sea in northern Germany. Developing such a prediction model is crucial for an AVM system with steering measures to enable more sustainable tourism and mobility. We contribute to both theory and practice by evaluating different prediction models concerning their accuracy and elaborating about how this prediction can be used as an effective countermeasure to overcrowding and unsustainable tourism.

2 Related Work

Research in tourism occupancy prediction has grown rapidly since 2006 [Li19]. In the following, we provide an overview considering the perspectives of (i) data sources, (ii) prediction models, and (iii) spatiotemporal granularity.

Data Sources: Traditionally, occupancy prediction was conducted with one-dimensional

historical data capturing a certain timespan. Such data can be collected manually (e.g., based on surveys) or automatically (e.g., based on cameras, smartphones, or advanced sensors) [APB21]. Suitable data sources include tourist arrivals (e.g., entries to national parks [Ab21] or countries [Ki21]), historical parking data [Ch19], web traffic [PY17], booking data (e.g., historical room allocation in hotels [PS21], [Zh18]), and payment data [APB21]. In order to improve predictions, several studies added supplementary information, such as hotel room prices [TB20], as well as weather and holiday information [Bi21]. Considering the importance of online behavior of potential visitors [GV20], recent research addresses the use of behavioral online data [WSS17], such as search engines and Google trends [Vo19], [Di19], [BL19], [Fe19], online reviews [Hu22], Facebook [ÖGG20], and sentiments [ÖGS19].

Prediction Models: Previous research [SQP19], [JC19], [WSS17] classifies occupancy prediction algorithms in the three categories (i) time-series models, (ii) econometric and statistical models, and (iii) ML-based models. Time series models remain the most widely used techniques [MR16], although ML-based technologies have proliferated significantly in recent years. However, ML models such as gradient boosting with XGBoost, which have shown promising results in other applications like taxi demand [Va18], bus demand [SSD21], or parking [APB21], are still rarely used in POI occupancy prediction.

Spatiotemporal Granularity: Regarding time and space, occupancy predictions address different levels of granularity. Considering their temporal resolution, existing approaches address seasonal [Ch15], to monthly [BL19], weekly [Zh18], [PY17], daily [TB20], [Bi21], [Ki21], [PS21], or even hourly [ZHL21] prediction. Prilistya et al. [PEF20] observed that monthly granularity is the most widely used data frequency in tourism occupancy prediction, but due to the more widespread use of ML, Jiao and Chen [JC19] observed a trend towards smaller time granularities. In terms of spatial resolution, approaches range from predicting tourist arrivals in countries [Ki21], regions [YZ19], or cities [Cl20], [TB20] to predicting occupancy of smaller areas. However, small areas usually do not refer to open or semi-open POIs like a beach, but rather address closed environments with well-defined entrances, such as parking lots [APB21], [Ch19], [ZZ20] or hotel rooms [PS21], [Zh18].

Overall, previous research rarely focuses on semi-open POIs to predict overcrowding situations in an AVM system [Kh20]. Further, the literature review indicates advanced ML models like XGBoost are not yet applied and compared with other approaches to predict touristic POI occupancy. Our work relies on real-world sensor data to predict POI occupancy of an exemplary German destination with a fine-grained temporal prediction up to ten days in advance. The comparison of a classical time series model with various ML models of different types illustrates the use case specific usability of each model. We discuss the results of the models regarding their application as a countermeasure to overcrowding and unsustainable tourism.

3 Methodology and Case Study

The methodology to develop an occupancy prediction model is based on the Data Science Trajectories (DST) map by Martinze-Plumed et al. [Ma19], which is a more sophisticated version of the classic CRISP-DM approach [Ch00]. The DST-map is a more flexible approach to data science projects that integrates further steps regarding exploratory and data management activities around the CRISP-DM model.

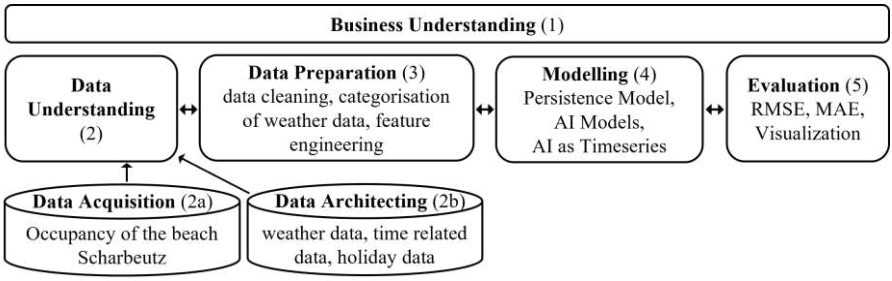


Fig. 1: The DST Map illustrates all steps within this study from business understanding to evaluation

Fig. 1 presents the DST map with the individual adapted steps for this study. In the Business Understanding part, we first provide an overview of the case study objective, followed by the examination of the utilized data. We divide the data understanding part in the acquisition of occupancy data and the architecting of external influential data which are both prepared cyclically in steps two and three. Subsequently the fourth step presents different model types with their inherent ML models, followed by the evaluation approach.

3.1 Business understanding: Predicting the occupancy of the beach in Scharbeutz

To develop a prediction model for the occupancy of a POI, we conduct a case study of the Bay of Lübeck in Scharbeutz at the Baltic Sea in northern Germany. The Baltic Sea is the most popular tourist destination of the German population with a continuously increasing trend [Ar22]. This trend was reinforced by the increase in domestic travel due to the Covid-19 pandemic [In21]. The high proportion of day visitors, accounting for around $^2/_3$ of the total visitors, has the effect that relatively scarce information is available on occupancy leading to unpredictable peaks in beach occupancy [De19], [De16]. To close the existing information gap, sensors have been installed at the entrances to the beach to count people entering and leaving the beach. The current state of data utilization is the representation of the current occupancy of the beach, which does not yet include a future prediction [To22]. However, for an active visitor management, a prediction, especially about occupancy peaks, is necessary to initiate possible steering measures in advance. Steering measures can be roughly divided into short-term measures like for example the selection of another parking lot and medium-term measures such as choosing another POI. Therefore, the different objectives of steering measures require different prediction spans

and accuracy. In the following case study, we use the collected occupancy data of the beach in Scharbeutz to develop, evaluate, and discuss the applicability and performance of different prediction models.

3.2 Data understanding and data preparation

The data acquisition phase covers the collecting of occupancy data of the beach in Scharbeutz which is also available as open data [Mi22]. The raw data contains quarter-hourly summed numbers of people entering and leaving the beach over the entire Scharbeutz beach section from 19-08-2020 to 23-03-2022. This sum of people entering and leaving is intended to represent the occupancy of the beach. A first analysis reveals annual seasonality with higher occupancy in the summer months and a daily seasonality between day and night. Further, some days demonstrate a high and irregular peak occupancy. Since the entrances of the adjoining beach sections are not equipped with sensors, the absolute number of people at the beach is probably not exact, but the general trends and peak occupancies can be identified. Data preparation of the occupancy data as the target value includes elimination of missing values and aggregation to larger time periods. To eliminate gaps in the time series due to missing data, we created a continuous time series dataframe and filled the small amount of missing data with zeros. Since the fine-grained partitioning into 15-minute time periods cause increased noise, we aggregated the data to larger 6 hour time periods per day, starting at midnight (see Tab. A.1).

The data architecting phase covers researching and compiling of possible factors influencing the target variable of occupancy. We identified three areas of possible influencing factors: time-related features, holiday-related features, and weather-related features. The features' data are required historically and in the near future, as historic data is necessary to train and test the prediction models, whereas future data is required for a real-world application of the developed prediction model. The time-related features are created to include and identify the possible timing of seasonality in the prediction model as an input variable. The holiday-related features are included as a possible reason for the peak demand since the population's leisure time should have an influence on the beach occupancy. Due to the high proportion of day visitors who presumably live near the considered POI, school holiday of Schleswig-Holstein is a specific feature. Since the Baltic Sea is a popular destination for domestic travel, the general school holiday density in Germany is covered as a feature to may represent increased occupancy by overnight visitors [Ar22]. Since visiting a beach is an outdoor-only activity, we assume that the occupancy is highly dependent on the weather. The collected weather data is from the open data hub of the German Weather Service which provides historical data as well as weather forecasts up to 10 days [Ge22b]. Similar to the occupancy data, we aggregated the raw weather data to 6-hour time periods. Following Studer et al. [St21], we converted the continuous raw data values into categorical features to ensure a good performance of the prediction models and to ensure that we use the same input features for each model. The categorization of the weather data follows the standards of the German Weather Service such as the classification of temperature, precipitation intensity, and wind class

[Ge22c], [Ge22a], [Ge22d] (see Tab. A.2). Precipitation form is not directly available for the historical and future weather data which is why we created the feature via feature engineering by combining input variables (see Tab. A.3). After categorizing the four weather-related features of wind, rain, temperature, and precipitation form, we transformed the features from categorical to binary features using one-hot encoding [St21]. Tab. 1 illustrates all input features, for the structure of the final input dataset see Tab. A.4.

Time-related	**Holiday**	**Weather**
year	weekend	wind category
month	bridging day	rain category
day of year	public holiday (bank holiday)	temperature category
quarter of year	regional school holiday	precipitation form
day of month	German school holiday density	
calendar week		
day of week		
hour		

Tab. 1: Overview of the features after categorization and feature engineering

3.3 Modelling

In this case study, we aim to identify a model for predicting the occupancy of the beach in Scharbeutz and therefore consider four different model types. The first model type (1) is a simple persistence model whose results are used as a baseline. It simply takes a historical value with a predefined lag as the prediction. To obtain a reasonably good baseline, we tested lags from one time step up to 1460 time steps, which equals exactly one year. The second type are supervised ML models (2) that aim to predict the target variable y, here beach occupancy per time period, based on several input features X [JM15]. Tab. 2 illustrates a brief overview of the ML models. We implemented the ML models in Python using the scikit-learn packages, and we avoided excessive tuning of the models to allow for an even comparison The third model type are ML time series models (3), which merely consider the historical occupancy data as an input variable for prediction [Ki10]. For comparison, we use the same supervised ML models as in the second model type. To apply ML models as time series prediction models, we first transformed the time series into a supervised learning model. We created the required input features by shifting the occupancy data backwards by a predefined lag. To improve the prediction quality, 10 input features were created by backward shifting the data with a lag from 1 to 10. The target variable y remains the original occupancy of the beach data. Prediction takes place with a walk forward validation, where exactly one time step ahead is predicted, and the observed values are added to the training data set over time. The fourth model type is a classical approach (4) with the statistical, Seasonal Autoregressive Integrated Moving Average model (SARIMA$(p, d, q)(P, D, Q, m)$) [HX98], [LI10]. The parameters of the SARIMA model were determined using the automatic optimization of pmdarima with the seasonal

parameter set to one day corresponding to four timesteps, resulting in final parameters of $(0,1,1)(0,1,1,4)$ [Sm22]. Analogous to the ML models as time series, we performed the prediction with SARIMA with a walk forward validation.

ML model	Short description	Source
Extreme Gradient Boosting (XGBoost)	The XGBoost model is a system optimized implementation of gradient boosting, whereby we included early stopping to avoid overfitting. Weak learners of decision trees are gradually improved and merged to obtain a strong one as the final prediction model.	[Fr01], [xg22]
Random Forest	The Random Forest model generates various independent trees by randomly selecting a choice of features and averaging over the results for prediction. Here, a maximum depth of four levels was used.	[Al12], [sc22a]
Support Vector Regression (SVR)	The SVR model creates an inherent function for predictions based on a kernel function by penalizing extreme values that are not within a specified radius. By testing the different kernels, the linear kernel indicated the best overall results here.	[AK15], [Va00], [sc22b]

Tab. 2: Overview of the considered ML models

3.4 Evaluation

To evaluate the performance of the different models, we split the occupancy data into a training and a testing data set, whereby the training data set includes the target variable to fit the model. With the test data set, the fitted model generates predictions which are then compared to the expected target values. In this case study, we split the time series data on the 1st of November 2021 which results in a training data set of 74% of the overall data set and a testing data set of 26%. With the single split, we aim to enable an even comparison of the different models and model types. To compare the predicting performance of the various models, we used the performance metrics root mean squared error ($RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$) and mean absolute error ($MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$). However, since in this case study especially the prediction of days with peak occupancy rates are relevant, a graphical interpretation is included as well.

4 Results

Tab. 3 illustrates the performance metrics RMSE and MAE for the testing dataset of the case study for each developed model and model type. For each model type (1)-(4) the best performance of RMSE and MAE is highlighted green. If a model performed worse than the best performance of the persistence model, the value of the performance metric is

highlighted red. The persistence model type (1) indicated the best results with a lag of one or four timesteps, whereby the lag of 1460 time steps, which corresponds to exactly one year, being the worst. This effect may be explainable by the limited amount of data which was mainly collected during the Covid-19 pandemic with many unusual effects. Thereby a smaller lag of 1 or 4 timesteps led to better results. For the ML models (2), the XGBoost model indicated the best results in both metrics. The Random Forest model performed the worst and was even below the baseline of the persistence model in the metric MAE. For the ML models as time series (3) XGBoost and Random Forest were particularly close and performed best across all model types. In comparison, the SVR models performed nearly the same in the model types (2) and (3). Due to the walk-forward approach by predicting one time step after another, all of the ML time series models (3) took considerably longer than the ML models (2). Since the Random Forest model takes a relatively long time to fit the model, it was the slowest model. The SARIMA model (4) performed only slightly better than the simple persistence model in the RMSE, yet worse in the MAE measurement. Since SARIMA is a statistical model, the prediction time was relatively fast and comparable to ML models (2) despite the walk-forward approach.

Model Type	**Models**		
(1) Persistence Model	lag = 1 (6 hours)	lag = 4 (1 day)	lag = 1460 (1 year)
RMSE	161	162	285
MAE	93	79	127
(2) ML Models	XGBoost	Random Forest	SVR
RMSE	117	140	137
MAE	60	94	80
(3) ML as Time series	XGBoost	Random Forest	SVR
RMSE	96	98	128
MAE	51	46	64
(4) SARIMA Time series Model			
RMSE	130		
MAE	82		

Tab. 3: RMSE and MAE of the prediction models of all model types

For an interpretation of the performance metrics, Fig. 2 illustrates the predicted and expected values for November 2021 for the model types (2), (3), and (4). The XGBoost model and the Random Forest model demonstrate similar effects in both types (2) and (3). As ML models (2), XGBoost and Random Forest demonstrate a continuous, recurring daily trend, but successfully identify days with peak occupancy. The difference in the performance metrics is explainable as the XGBoost model differentiates between low and medium occupancy, whereas the Random Forest model predicts the same relatively high daily trend resulting in poorer performance metrics. The ML models as time series (3), demonstrate exceptionally good predictions in November 2021, with both models correctly identifying peak occupancy days as well as lower occupancy days, and only slightly exceeding the expected values. The still relatively high error rates are related to days such as New Year's Day, which display a strongly deviating pattern from 2020/21 to

2021/22 due to the Covid-19 pandemic restrictions that are not yet covered in the trained model. The poor performance of the SVR model in both model types (2) and (3) is immediately apparent, as only the slight daily trend is anticipated, but it detected no occupancy peaks. The penalization of extreme values by creating the internal function of the model probably causes this effect. For predicting days with occupancy peaks, this model seems unsuitable. The SARIMA model correctly predicts the seasonal trend between day and night but does not identify days with peak occupancy in advance. The walk-forward validation only anticipates the upward or downward trend over time.

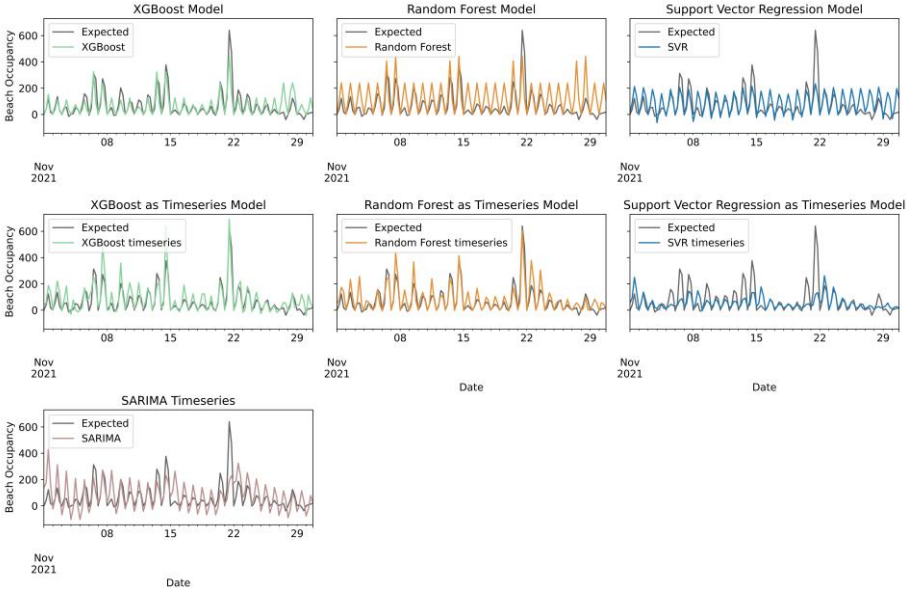


Fig. 2: Predicted versus expected occupancy data for November 2021 of all prediction models

5 Discussion

The performance metrics and graphical interpretation indicate a preference for the time series models with XGBoost or Random Forest. However, the time series models predict only one time step ahead, which is equal to 6 hours. In contrast, the supervised ML models predict 40 time steps which is equal to 10 days and limited only by the weather forecast as an input parameter. Since this case study aims to develop an occupancy prediction model to enable active management of beach visitors in Scharbeutz, the prediction of only 6 hours in advance is probably too short. Steering measures to avoid peak occupancies at the beach are more likely to be medium-term to convince some visitors to choose another POI. In this scenario, the still quite good performance of XGBoost as supervised ML model is in favor of the slightly more accurate ML time series models.

Further, the performance of the XGBoost model may improve with an increasing amount of available data. For example, the special situation of the Covid-19 pandemic and its resulting restrictions should become relative as more data with a regular history become available. In this context, a feature that indicates the presence of restrictions like contact restrictions and curfew gives the model more information about the special situation at New Year 2020/21. Since no data was available in this case study, the mentioned Covid-19-specific feature would not have influenced the predictive performance as no comparison to non-pandemic years was possible. Regardless of the Covid-19 pandemic, integrating additional features like events at the beach, felt weather temperature, or water temperature might improve model performance.

Overall, the selection of prediction models itself highly depends on the use case. On the one hand, if the prediction is utilized in steering measures to avoid overcrowding, like in this use case in Scharbeutz, a medium- to long-term prediction is required, since visitors need to be addressed already in the planning phase of their trip. As a countermeasure to overcrowding, the predicted occupancy information can be utilized in an AVM system by recommending alternative POIs to potential visitors. Here, minor deviations in the prediction are not decisive since there is no specific and limiting occupancy number for an open POI like a beach. On the other hand, steering measures at the POI, such as a controlled filling of parking lots requires a short-term and accurate prediction.

6 Conclusion

In the present paper, we elaborate on the applicability and performance of various model types, including different ML models, to predict the occupancy of tourist destinations. In a case study of beach occupancy in Scharbeutz, northern Germany, the time series models with XGBoost and Random Forest demonstrated the lowest error, followed by the supervised XGBoost model. However, due to the limited prediction of only one time step in advance, the time series prediction is quite limited and only suitable for steering visitors to a limited extent. In such cases, the supervised XGBoost model may be in favour since peak occupancy days were still predicted correctly. Overall, the case study highlights the importance of selecting the right model depending on the use case and the associated objectives. In future research, we plan to integrate more features on a larger dataset to improve the performance of the ML model, and plan to evaluate it with relative measures and time series cross-validation to obtain a more reliable estimate. Subsequently, the most important features might be identified by using SHAP values, which would allow explaining the underlying effects of peak occupancy days. In addition, the model can be further validated by integrating parking data near the beach to verify whether a full parking lot indicates a high beach occupancy. Finally, the transferability of the developed models to other regions and POIs, such as the Allgäu region in the south of Germany, might be investigated.

7 Acknowledgements

The authors gratefully acknowledge the financial support of the Project “AIR” (67KI21005G) by the Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection of Germany (BMUV) and the financial support of the Project “FEB-NAFV” (19F2198A) by the Federal Ministry for Digital and Transport (BMDV) as part of the mFUND innovation initiative.

Bibliography

[Ab21] Abu, N. et al.: SARIMA and Exponential Smoothing model for forecasting ecotourism demand: A case study in National Park Kuala Tahan, Pahang. Journal of Physics: Conference Series 1/1988, pp. 12118, 2021.

[AK15] Awad, M.; Khanna, R.: Support Vector Regression. In (Awad, M.; Khanna, R. Eds.): Efficient Learning Machines. Apress, Berkeley, CA, pp. 67–80, 2015.

[Al12] Ali, J. et al.: Random forests and decision trees. International Journal of Computer Science Issues (IJCSI) 9, pp. 272, 2012.

[APB21] Assemi, B.; Paz, A.; Baker, D.: On-Street Parking Occupancy Inference Based on Payment Transactions. IEEE Transactions on Intelligent Transportation Systems, pp. 1–12, 2021.

[Ar22] Arbeitsgemeinschaft Verbrauchs- und Medienanalyse (VuMA): Verbrauchs- und Medienanalyse - VuMA 2022. Freizeit, Urlaub, Reisen, 2022.

[Bi21] Bi, J.-W. et al.: Forecasting Daily Tourism Demand for Tourist Attractions with Big Data: An Ensemble Deep Learning Method. Journal of Travel Research, 2021.

[BL19] Bokelmann, B.; Lessmann, S.: Spurious patterns in Google Trends data - An analysis of the effects on tourism demand forecasting in Germany. Tourism Management 75, pp. 1–12, 2019.

[Ca19] Capocchi, A. et al.: Overtourism: A Literature Review to Assess Implications and Future Perspectives. Sustainability 12/11, p. 3303, 2019.

[Ch00] Chapman, P. et al.: CRISP-DM 1.0 step-by-step data mining guide, 2000.

[Ch15] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Chen, K.: Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), pp. 1–4, 2015.

[Ch19] Chawathe, S. S.: Using Historical Data to Predict Parking Occupancy: 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE, New York City, NY, USA, pp. 534–540, 2019.

[Cl20] Claude, U.: Predicting Tourism Demands by Google Trends: A Hidden Markov Models Based Study. Journal of System and Management Sciences, 2020.

[De16] Deutschen Wirtschaftswissenschaftlichen Institut für Fremdenverkehr e. V.: Wirtschaftsfaktor Tourismus für das Reisegebiet Ostsee (Schleswig-Holstein), 2016.

[De19] Deutschen Wirtschaftswissenschaftlichen Institut für Fremdenverkehr e. V. (dwif): Wirtschaftsfaktor Tourismus für das Reisegebiet Ostsee 2019, 2019.

[Di19] Dinis, G. et al.: Google Trends in tourism and hospitality research: a systematic literature review. Journal of Hospitality and Tourism Technology 4/10, pp. 747–763, 2019.

[Ei22] Eisele, J. et al.: Besucherlenkung und Reduktion des motorisierten Freizeitverkehrs - das Potential datengetriebener und flexibler Busangebote. In (Teich, T. et al. eds.): Innovation und Kooperation auf dem Weg zur All Electric Society. Emergenzen für neue Geschäftsprozesse. Springer, Heidelberg, 2022.

[Fe19] Feng, Y. et al.: Forecasting the number of inbound tourists with Google Trends. Procedia Computer Science 162, pp. 628–633, 2019.

[Fr01] Friedman, J. H.: Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 5/29, pp. 1189–1232, 2001.

[Ge22a] German Weather Service: Weather and climate - Deutscher Wetterdienst - Glossary - N - Precipitation intensity. https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv2=101812&lv3=101906, accessed 17 Mar 2022.

[Ge22b] German Weather Service: Climate Data Center (CDC). https://opendata.dwd.de/climate_environment/CDC/, accessed 17 Mar 2022.

[Ge22c] German Weather Service: Formulations of the Weather Elements. https://www.dwd.de/DE/service/lexikon/begriffe/W/Wetterelementeformulierungen_pdf.pdf?__blob=publicationFile&v=3, accessed 17 Mar 2022.

[Ge22d] German Weather Service: Weather and climate - German Meteorological Service - Glossary - B - Beaufort scale. https://www.dwd.de/DE/service/lexikon/Functions/glossar.html?lv2=100310&lv3=100390, accessed 17 Mar 2022.

[GG18] Gross, S.; Grimm, B.: Sustainable mode of transport choices at the destination – public transport at German destinations. Tourism Review 3/73, pp. 401–420, 2018.

[GV20] Gowreesunkar, V. G.; Vo Thanh, T.: Between Overtourism and Under-Tourism: Impacts, Implications, and Probable Solutions. In (Séraphin, H.; Gladkikh, T.; Vo Thanh, T. eds.): Overtourism. Springer International Publishing, Cham, pp. 45–68, 2020.

[Hø00] Høyer, K. G.: Sustainable Tourism or Sustainable Mobility? The Norwegian Case. Journal of Sustainable Tourism 2/8, pp. 147–160, 2000.

[Hu22] Hu, M. et al.: Tourism demand forecasting using tourist-generated online review data. Tourism Management 90, p. 104490, 2022.

[HX98] Ho, S. L.; Xie, M.: The use of ARIMA models for reliability forecasting and analysis. Computers & Industrial Engineering 1-2/35, pp. 213–216, 1998.

[In21] Institut für Demoskopie Allensbach: Allensbacher Markt- und Werbeträger-Analyse - AWA 2021. Urlaub und Reisen, 2021.

[JC19] Jiao, E. X.; Chen, J. L.: Tourism forecasting: A review of methodological developments over the last decade. Tourism Economics 3/25, pp. 469–492, 2019.

[JM15] Jordan, M. I.; Mitchell, T. M.: Machine learning: Trends, perspectives, and prospects. Science (New York, N.Y.) 6245/349, pp. 255–260, 2015.

[Kh20] Khatibi, A. et al.: Fine-grained tourism prediction: Impact of social and environmental features. Information Processing & Management 2/57, p. 102057, 2020.

[Ki10] Kitagawa, G.: Introduction to time series modeling. Chapman & Hall/CRC, Boca Raton, 2010.

[Ki21] Kim, D.-K. et al.: A Daily Tourism Demand Prediction Framework Based on Multi-head Attention CNN: The Case of The Foreign Entrant in South Korea: 2021 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, Orlando, FL, USA, pp. 1–10, 2021.

[LI10] Loganathan, N.; Ibrahim, Y.: Forecasting international tourism demand in Malaysia using Box Jenkins Sarima application. South Asian Journal of Tourism and Heritage 2/3, pp. 50–60, 2010.

[Li19] Liu, H. et al.: Hot topics and emerging trends in tourism forecasting research: A scientometric review. Tourism Economics 3/25, pp. 448–468, 2019.

[Ma19] Martinez-Plumed, F. et al.: CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. IEEE Transactions on Knowledge and Data Engineering 8/33, pp. 3048–3061, 2019.

[Mi22] Ministerium für Energiewende, Landwirtschaft, Umwelt, Natur und Digitalisierung: Datensätze - Open-Data Schleswig-Holstein. https://opendata.schleswig-holstein.de/dataset?q=&ext_startdate=&ext_enddate=&groups=tran&sort=score+desc%2C+metadata_modified+desc, accessed 12 Apr 2022.

[Ne22] Neubig, S. et al.: Data-driven Initiatives of Destinations Supporting Sustainable Tourism: Americas Conference on Information Systems (AMCIS) 2022, 2022.

[ÖGG20] Önder, I.; Gunter, U.; Gindl, S.: Utilizing Facebook Statistics in Tourism Demand Modeling and Destination Marketing. Journal of Travel Research 2/59, pp. 195–208, 2020.

[ÖGS19] Önder, I.; Gunter, U.; Scharl, A.: Forecasting Tourist Arrivals with the Help of Web Sentiment: A Mixed-frequency Modeling Approach for Big Data. Tourism Analysis 4/24, pp. 437–452, 2019.

[Pa22] Paidi, V. et al.: CO2 Emissions Induced by Vehicles Cruising for Empty Parking Spaces in an Open Parking Lot. Sustainability 7/14, p. 3742, 2022.

[PEF20] Prilistya, S. K.; Erna Permanasari, A.; Fauziati, S.: Tourism Demand Time Series Forecasting: A Systematic Literature Review: 2020 12th International Conference on Information Technology and Electrical Engineering (ICITEE). IEEE, Yogyakarta, Indonesia, pp. 156–161, 2020.

[PS21] Phumchusri, N.; Suwatanapongched, P.: Forecasting hotel daily room demand with transformed data using time series methods. Journal of Revenue and Pricing Management, 2021.

[PY17] Pan, B.; Yang, Y.: Forecasting Destination Weekly Hotel Occupancy with Big Data. Journal of Travel Research 7/56, pp. 957–970, 2017.

[sc22a] scikit-learn developeres: sklearn RandomForestRegressor. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html, accessed 24 Apr 2022.

[sc22b] scikit-learn developeres: sklearn svm.SVR. https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVR.html, accessed 24 Apr 2022.

[Sc22] Schmücker, D. et al.: Digitales Besuchermanagement im Tourismus. Konzeptioneller Rahmen und Gestaltungsmöglichkeiten. In (Gardini, M. A.; Sommer, G. Eds.): Digital Leadership im Tourismus. Digitalisierung und Künstliche Intelligenz als Wettbewerbsfaktoren der Zukunft. Springer, Wiesbaden, 2022.

[Sm22] Smith, T. G.: pmdarima 1.8.5 documentation. pmdarima.arima.auto_arima. https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html, accessed 24 Apr 2022.

[SQP19] Song, H.; Qiu, R. T.; Park, J.: A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. Annals of Tourism Research 75, pp. 338–362, 2019.

[SSD21] Stadler, T.; Sarkar, A.; Dünnweber, J.: Bus Demand Forecasting for Rural Areas Using XGBoost and Random Forest Algorithm. In (Saeed, K.; Dvorský, J. Eds.): Computer Information Systems and Industrial Management. Springer International Publishing, Cham, pp. 442–453, 2021.

[St21] Studer, S. et al.: Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology. Machine Learning and Knowledge Extraction 2/3, pp. 392–413, 2021.

[TB20] Tsang, W. K.; Benoit, D. F.: Gaussian processes for daily demand prediction in tourism planning. Journal of Forecasting 3/39, pp. 551–568, 2020.

[To22] Tourismus-Agentur Lübecker Bucht: Strandticker Lübecker Bucht. https://www.luebecker-bucht-ostsee.de/strandticker, accessed 12 Apr 2022.

[Va00] Vapnik, V. N.: The Nature of Statistical Learning Theory. Springer, New York, NY, 2000.

[Va18] Vanichrujee, U. et al.: Taxi Demand Prediction using Ensemble Model Based on RNNs and XGBOOST: 2018 International Conference on Embedded Systems and Intelligent Technology & International Conference on Information and Communication Technology for Embedded Systems (ICESIT-ICICTES). IEEE, pp. 1–6, 2018.

[Vo19] Volchek, K. et al.: Forecasting tourist arrivals at attractions: Search engine empowered methodologies. Tourism Economics 3/25, pp. 425–447, 2019.

[Wo18] World Tourism Organization et al.: 'Overtourism'? - Understanding and Managing Urban Tourism Growth beyond Perceptions, Executive Summary. UNWTO, 2018.

[WSS17] Wu, D. C.; Song, H.; Shen, S.: New developments in tourism and hotel demand modeling and forecasting, p. 23, 2017.

[xg22] xgboost developers: XGBoost Documentation — xgboost 1.5.2 documentation. https://xgboost.readthedocs.io/en/stable/index.html, accessed 13 Apr 2022.

[YZ19] Yang, Y.; Zhang, H.: Spatial-temporal forecasting of tourism demand. Annals of Tourism Research 75, pp. 106–119, 2019.

[Zh18] Zhang, M. et al.: Weekly Hotel Occupancy Forecasting of a Tourism Destination. Sustainability 12/10, p. 4351, 2018.

[ZHL21] Zheng, W.; Huang, L.; Lin, Z.: Multi-attraction, hourly tourism demand forecasting. Annals of Tourism Research 90, p. 103271, 2021.

[ZZ20] Zhao, Z.; Zhang, Y.: A Comparative Study of Parking Occupancy Prediction Methods considering Parking Type and Parking Scale. Journal of Advanced Transportation 2020, pp. 1–12, 2020.

Appendix

Part of the day	**Lower Limit**	**Upper Limit**
0: night	$\geq 12{:}00\ AM$	$< 6{:}00\ AM$
6: morning	$\geq 6{:}00\ AM$	$< 12{:}00\ PM$
12: afternoon	$\geq 12{:}00\ PM$	$< 6{:}00\ PM$
18: evening	$\geq 6{:}00\ PM$	$< 12{:}00\ AM$

Tab. A.1: 6-hour time periods

Temp. category	Lower Limit [°C]	Upper Limit [°C]	**Rain category**	Lower Limit $[kg/m^2]$	Upper Limit $[kg/m^2]$	**Wind category**	Lower Limit $[m/s]$	Upper Limit $[m/s]$
-10		< -10	0.5		< 0.5	0		< 0.3
0	≥ -10	< 0	2.5	≥ 0.5	< 2.5	1	≥ 0.3	< 1.6
10	≥ 0	< 10	5.0	≥ 2.5	< 5.0	2	≥ 1.6	< 3.4
18	≥ 10	< 18	10.0	≥ 5.0	< 10.0	3	≥ 3.4	< 5.5
24	≥ 18	< 24	50.0	≥ 10.0	< 50.0	4	≥ 5.5	< 8.0
30	≥ 24	< 30	51.0	≥ 50.0		5	≥ 8.0	< 10.8
35	≥ 30	< 35				6	≥ 10.8	< 13.9
36	≥ 35					7	≥ 13.9	< 17.2
						8	≥ 17.2	< 20.8
						9	≥ 20.8	< 24.4
						10	≥ 24.4	< 28.4
						11	≥ 28.4	< 32.6
						12	≥ 32.6	

Tab. A.2: Temperature, rain, and wind categories

Precipitation form	Historical observation precipitation form	Mosmix forecast
0: no precipitation	NaN, 0, 4, 9	Snow water equivalent = 0 & Precipitation height < 0.5
1: rain	1, 6	Precipitation height > 0.5
2: snow	7, 8	Snow water equivalent > 0

Tab. A.3: Feature engineering of the precipitation form

Features	Type of Data	Value Ranges
year	Integer	[2020, 2021, 2022]
month	Integer	[1, …, 12]
day of year	Integer	[1, …, 365]
quarter of year	Integer	[1, …, 4]
day of month	Integer	[1, …, 31]
calendar week	Integer	[1, …, 52]
day of week	Integer	[1, …, 7]
hour	Integer	[0, …., 23]
weekend	Binary	[0, 1]
bridging day	Binary	[0, 1]
public holiday (bank holiday)	Binary	[0, 1]
regional school holiday	Binary	[0, 1]
German school holiday density	Float	[0, …, 1]
wind category	categorical	[0, 1, 2, 3, 4, 5, 6, 7,8, 9, 10, 11, 12]
rain category	categorical	[0.5, 2.5, 5.0, 10.0, 50.0, 51.0]
temperature category	categorical	[-10, 0, 10, 18, 24, 30, 35, 36]
precipitation form	categorical	[0, 1, 2]

Tab. A.4: Structure of the final data set of input features

Aligning Value Creation in Ecosystems through Roadmapping

Dimitri Petrik [1], Stefan Trieflinger[2] and Felix Schönhofen[1]

Abstract: There is a growing consensus in research and practice that value-creating networks and ecosystems are supplementing the traditional distinction between the internal firm and market perspectives. To achieve joint value in ecosystems, it is crucial to align the various interests of independently acting ecosystem actors and create a common vision. In this paper, we argue that the ecosystem-wide use of product roadmaps may help with this. To get a better understanding of how roadmapping is conducted in the dynamic ecosystem environment, we systematize the main characteristics of product roadmaps and perform a conceptual comparison with the known challenges of ecosystem management. Comparing the two concepts of ecosystems and product roadmaps, we highlight the fit between the characteristics and objectives of the roadmaps and the challenges of ecosystem management. Hence, we propose to experiment with the ecosystem-wide use of product roadmaps as well as the empirical study of the challenges emerging in the process and the associated redesign of the roadmaps.

Keywords: Ecosystems, Value Networks, Ecosystem Management, Ecosystem Orchestration, Value Alignment, Product Roadmaps, Roadmapping, Ecosystem Roadmaps.

1 Uncertainties Navigating Ecosystems and Dynamic Networks

Digitalization rapidly changes the market conditions for business organizations, reshaping the product design and competition rules. The integration of information technologies, for instance, makes products more intelligent and networked [PH14], [SHL20]. Digitalization even fosters the decoupling of information, functionality, and technologies from the physical products, blurring the traditional organizational boundaries and fostering systemic innovation [LN15], [SHL20]. However, many digitalization initiatives did not realize their potential by relying on their own capabilities despite the changes in product design described earlier. Whether data or digital infrastructure, it becomes evident that companies can only realize the superior value of digital services in ecosystems by balancing their internal resources with the external resources [Sk19], [JSP21].

[1] University of Stuttgart, Dept. VIII: Information Systems II,, 70174, Stuttgart, Germany
dimitri.petrik@bwi.uni-stuttgart.de, https://orcid.org/0000-0002-0244-1235
[2] Hochschule Reutlingen, Computer Science, Alteburgstraße 150, 72762 Reutlingen,
stefan.trieflinger@reutlingen-university.de
[3] University of Stuttgart, Dept. VIII: Information Systems II,, 70174, Stuttgart, Germany
felix.schoenhofen@bwi.uni-stuttgart.de

doi:10.18420/inf2022_35

Ecosystems represent dynamic cooperation forms to organize value creation, particularly effective in sourcing external resources and capabilities to achieve efficient value creation [TA18], [DAA18]. Ecosystems also build a structure to organize a set of multilateral firms to contribute to a shared value proposition [Ad17]. Accordingly, ecosystem actors are independent firms who may collaborate to jointly realize a specific value manifestation and compete to realize other products or services. The presence of coopetition underlies the prevalent dynamics and fuels a higher degree of uncertainty during the value creation in ecosystems. To illustrate this source of uncertainty, it is worth taking a closer look at platform-based ecosystems, an ecosystem type where ecosystem actors usually agree to utilize a digital platform. Platform-based ecosystems can suffer from information and power asymmetries since the platform provider has control over the fundamental ecosystem building block [He20]. If a platform provider, as a key technology supplier, decides to change its strategy or architecture (e.g., depreciation of Flash by Apple for iOS), this may significantly impact the product development of all ecosystem actors [Ea15]. The prevailing power asymmetry increases uncertainty in the value co-creation in ecosystems, as the value creation partners lack knowledge about the decisions of other ecosystem actors [HHD21]. Hence, ecosystem actors may strive to achieve different strategic goals, preventing beneficial cooperation in an ecosystem. Consequently, planning ahead without knowing what other ecosystem actors you depend on will do in the future is challenging and may even lead to an ecosystem demise, indicated by less value-creating activities.

One of the central sustainability approaches in networked business models is understanding that no sustainable value can be created for customers without creating value for a broader range of stakeholders [JTS20]. In the ecosystem context, we understand this economic sustainability approach as the necessity to manage a fair distribution of success among the value co-creating ecosystem actors. Against this background, ecosystems must establish fair governance and promote the shared value proposition of the ecosystem [TA18], [SP21]. Although hints can be found about the design importance of the ecosystem participation architecture [SP21], ecosystem management is still an infant discipline. Accordingly, there are gaps concerning prescriptive knowledge on how a functioning participation architecture can be designed concretely, while other known scientific disciplines of managing distributed value-creating networks, such as SCM, neglect critical ecosystem determinants [Go21]. Hence, in this paper, we argue that adapting the so-called roadmapping approach with the resulting roadmaps for ecosystem-wide use could work out as a potential solution to support success-critical ecosystem challenges, which may lead to ecosystem demise at some point. Product roadmaps are strategic communication tools that provide an overview of the future direction of product planning and include all important perspectives [Lo17], [Su11]. Product roadmaps align stakeholders' different perspectives and needs, aiming to gain their support through transparent coordination [Tr21]. Consequently, using product roadmaps to align the various ecosystem actors may be promising and motivates the following research question: *Which ecosystem challenges can be addressed through roadmapping*? Therefore, we intend to achieve a conceptual embedding of the roadmapping technique in the context of ecosystem management, reflecting its potential

against the known challenges in different ecosystem types. This enables a comparison of ecosystem challenges and roadmap potentials to shed light on how product roadmaps can support the management of dynamic networks such as ecosystems.

2 Ecosystem Value Creation

2.1 Background and definitions

As stated before, firms increasingly create value by cooperating with other firms in networks. Such networks can become dynamic, so Moore [Mo93] reinterpreted biological ecosystems to analyze the interdependence of firms during their interactions. Networks and ecosystems are cooperative forms of organization, with networks determined by dyadic relationships, and ecosystems determined by multilateral relationships [PS21]. Despite the similarities with the traditional research on networks of supply chains (i.e., exploring how organizations can manage their dependencies on the external environment), ecosystem research is considered to be in the definition stage [SG20], [Go21]. Prior research distinguishes at least six recognized ecosystem types [Gu20]. All the ecosystem types can be considered as structures of loosely coupled actors that cooperate and compete simultaneously, resulting in the actors' interdependency [Ad17], [JCG18]. Business ecosystems consist of multiple coopeting organizations to create a superior business value satisfying customer needs [Mo93], [Gu20]. Innovation ecosystems primarily refer to the co-creation of innovation output [Va15], [JCG18]. Service ecosystems organize co-created services [Sk19], [Gu20]. These ecosystem types are, in many cases, underpinned by a shared digital infrastructure (e.g., specific software or standards) to support exchange between ecosystem actors [Gu20]. If platform characteristics determine the commonly agreed digital infrastructure, it is more accurate to refer to platform-based ecosystems. Platform-based ecosystems are based on a central platform architecture serving as a technical foundation to realize value. So-called innovation platforms, for instance, consist of technological building blocks that platform providers and loosely coupled platform users share to create complementary products and services based on the platform [CGY19]. The platform-based output is understood as non-generic (i.e., unique or supermodular) complements since it interacts with the platform core and extends the platform functionality, constituting systemic innovation [PAJ17], [Go21].

2.2 Complexity of ecosystem structure and management

All of these ecosystem types share the complexity for decision-makers to navigate the firm, achieving optimal strategic positioning in an ecosystem [WN13], [CF21]. The complexity is partially caused by the distributed nature of ecosystemic value creation (understood as the *ecosystem structure*), as value-creating activities and the resulting non-generic complements are not directly controlled by a single firm [JCG18]. The structure

of ecosystem actors is complex, as **independent yet interdependent complementary firms** (complementors) co-create systemic innovation, which is usually not the case in value networks. Ecosystem actors act independently with no bilateral contractual basis, leading to additional competition for similar innovations. In addition, other ecosystem actors may be involved in the value creation process, such as suppliers, strategic partners, and end customers (e.g., contributing their data) across different domains [JCG18], [He20], [SG20]. The diversity of ecosystem actors can be high since ecosystems manifest organizational openness so that access to the ecosystem is open to potential enterprises outside the ecosystem. Under such conditions, ecosystems can scale more dynamically than traditional value networks and are therefore difficult to capture in long-term planning [Ga14], [He20], [PS21]. Hence, openness induces ecosystem dynamics, sacrificing the stability of closed innovation [WFC14], [CF21], adding new sources of uncertainty during the value creation activities. Ecosystem actors can, for instance, independently work on similar innovations in secret. Alternatively, ecosystem actors may sacrifice certain technologies or standards for the benefit of others. Therefore, ecosystem actors cannot be sure about the individual interests of other ecosystem actors, despite possible technological dependencies. Thus, decision-makers also associate ecosystems with the **risks of information asymmetries** for their firms. Besides, organizational openness may lead to an **oversupply of certain services** or decrease the quality if **"free-riders" join the ecosystem**. Empirical research has confirmed that uncontrolled ecosystem crowding with similar offerings fosters negative network effects, negatively impacting innovation incentives [Bo12]. Consequently, it is a challenge to **fully assess the value flows in ecosystems** to optimally incorporate one's own value contribution into the ecosystem. Accordingly, prior ecosystem research acknowledges the **necessity of careful ecosystem management** as a key to ecosystem success. In addition to the previously described trade-off related to ecosystem openness, orchestration becomes a critical capability for firms to successfully navigate in dynamic and complex environments such as ecosystems [BO18], [Go21].

Similar to the differentiation of different ecosystem types, current literature is also dominated by at least three terms, which conceptually overlap and indicate a conceptual ambiguity. These are ecosystem, orchestration, governance, and management. Ecosystem governance is a multidimensional discipline. Generally speaking, ecosystem governance sets which ecosystem actor (1) does what, (2) what does he control, and (3) how does he benefit [TA18], [DAA18], [TKB10]. Extant research has already discovered multiple variables of ecosystem governance [RS21]. To name a few recognized fields of action, the establishment of controlled access points, incentives for creating qualitative and unique complements, and distribution rights constitute the orchestration discipline. One of the critical action fields is the alignment of ecosystem actors for joint value creation without formal contract-based control. Therefore, orchestration is also defined as a construct aiming to define ecosystem roles, establish an appropriate participation architecture for filling these roles, and coordinate value-creating activities between the roles to achieve the ecosystem vision [DAA18], [Au21], [DP06]. Therefore, orchestration is about creating favorable conditions for valuable companies to actively contribute their resources, which

they only do if they achieve their individual goals to some degree [BO18]. A foundation of the ecosystem management concept covered by the literature categorizes it at a higher level. The latest research anticipates multiple views on ecosystem management. Managing the **interdependence tensions** or the complementor is assigned to the process view. Management of the ecosystem structure, rules, and orchestration are considered as the configurational view. Rearrangements of the ecosystem scope is an exemplary focus of the competitive view [Go21]. Tiwana also distinguishes between orchestration and governance in his book on platform-based ecosystems. From his point of view, governance actions that shape the architecture of the ecosystem (e.g., the underlying platform) enable and affect the orchestration of the ecosystem evolution, including the coordination of the ecosystem actors [Ti14]. It is noteworthy to add that governance can be either formal (i.e., in competitive ecosystems such as the iOS ecosystem) or informal (i.e., in collaborative ecosystems such as Wikipedia) [BO18]. However, the previously described openness in domains where value creation manifests in ecosystems, the **diversity of co-creating ecosystem actors** also reduces the chances of effective one-size-fits-all governance and significantly complicates the orchestration of stakeholders. Empirical examples such as the demise of the Symbian ecosystem, on the other hand, illustrate that excessive rigidity in governance does not meet the needs of ecosystem actors since a participation architecture must correspond to the ecosystem dynamics [WW14].

It has also been demonstrated that some companies could organize other firms around them, positioning themselves as hubs in ecosystem value-adding activities. Usually, such firms develop sufficient partner management capabilities or make a critical contribution to the ecosystem's core value proposition [IL04], [BO18], [JCG18]. On the one hand, this means that their value contribution can lie, among other things, in the provision of a digital platform that improves transactions between ecosystem actors or provides important technological building blocks for innovation. On the other hand, by controlling the digital infrastructure, hub companies can set standards and influence the evolution of platform-based complements and, in the best case, gain competitive advantages for the entire platform-based ecosystem. This strategy is also known as platform leadership [GC13]. However, platform leadership can also be considered as evidence of **power asymmetries** in ecosystems. Ecosystem actors may technically depend on the ecosystem infrastructure, and its significant changes may, as illustrated by the deprecation of Flash by Apple, it can jeopardize other actors' business models. Other platform providers exploit their position to replicate complements of other ecosystem actors [KBZ21]. Overall, the resulting power asymmetry is a constant risk in the delivery of ecosystem value. Accordingly, it is also assumed that especially incumbents from industrial domains are reserved about collaboration in ecosystems [CK21].

Digital platforms can coordinate activities, although the platform organization does not have to be a single firm [Ga14]. Accordingly, a hub organization is expected to exercise a certain degree of ecosystem governance over the ecosystem determinants, for instance, controlling the design of the underlying platform. In reality, governance does not have to reside with one company and can also be shared and open to other ecosystem actors [OK20]. Platform organizations are usually capable of setting the rules (1) of the

technological platform architecture, such as the modular core, standardized interfaces, and the complementary extensions and (2) social processes to guide the independent ecosystem actors [WFC14]. Platform governance consists of numerous instruments such as the **definition of decision rights, different controls of platform-based input, provision of boundary resources, strategic partnerships of the platform provider, preference of specific complements over others**, and much more [SWK16], [Go21], [RS21]. With this versatile set of measures, platform providers and ecosystem leaders can stimulate network effects, helping ecosystem actors achieve their goals while pursuing a shared ecosystem-wide vision without directly controlling them. It is assumed that other ecosystem actors will follow this direction and align around the ecosystem leaders [Mo96]. However, extant research did not offer support in the form of frameworks or tools, although each of the governance measures in itself represents a complex trade-off decision given the heterogeneous ecosystem structure. As a result of governing the ecosystem architecture or rule sets, orchestration may aim for different goals. It may satisfy the individual goals of innovative ecosystem actors, **correcting the power imbalances** among formally equal ecosystem actors, **fostering matchmaking** and innovative collaborations between actors, or **promoting specific complements** to attract certain customer groups. Accordingly, *trade-offs in governance and orchestration decisions* constitute one more ecosystem challenge category.

In addition to the uncertainties arising from ecosystem management decisions, the variety of *ecosystem dynamics* add to the ecosystem challenges and some examples of these are given below. Ecosystem actors may evolve into **multiple roles**, which can negatively affect innovation incentives if a certain ecosystem outcome becomes too much affected by competition [HW22], [Bo12]. Ecosystems may also be affected by so-called **bottlenecks**, which are components that prevent the overall positive development of the ecosystem. These may be critical technologies needed for a jointly created innovation or tendencies toward monopolistic behavior among orchestrating companies [HE17]. Furthermore, ecosystem actors may form **sub-ecosystems**, decoupling from the rest of the ecosystem, occurring as an unintended effect of orchestration measures. While sub-ecosystems can undoubtedly also generate unforeseen value (i.e., PhoneGap apps for mobile devices), there is also the potential for sub-ecosystems to jeopardize the ecosystem goals (i.e., performance deficits of PhoneGap apps). These ecosystem specifics can be addressed by ecosystem orchestration [JCG18] and thus require appropriate analysis techniques. Lastly, *competitive strategies* can contribute to ecosystem challenges. Known strategies such as **envelopment** or **platform injection** can also disrupt an ecosystem balance by exposing the ecosystem to competitive threats [SKD18], [KR20]. Therefore, applying such strategies should be based on systemic ecosystem analyses.

3 Roadmapping for Ecosystem Navigation

The previous section should outline why firms experience uncertainty, complexity, and high dynamics when navigating or orchestrating ecosystems. The following section

addresses the question of how roadmaps support coping with these inherent ecosystem challenges.

Roadmapping is a versatile technique that has been predominantly used in the context of individual companies as well as in the context of domains and markets for requirements engineering, strategic management, long-range planning, and other disciplines [Ka01], [AK03] [PFP04], [Su11], [Vi21]. Roadmapping is also a process of creating and revising roadmaps [LKK05], [Su11]. In their original form, roadmaps are graphic representations of the dynamic linkages between the three high-level strategy views on markets, products and technologies over time, enabling their exploration and communication to stakeholders [AK03], [Su11]. Diverse roadmap types can be combined as multiple layers and represent pathways from the actual state to a certain vision [PFP01], [PFP04], [AK03]. Roadmapping usually intends to create different forms of roadmaps, such as technology, product, innovation, project, or functions roadmaps [MI17]. In the last years, roadmaps were used to support transparent, business-driven, and cooperative decision-making during software development [VLR02], [LKK05]. One can observe that roadmaps have been used in different software development processes. However, recent research has found that feature-based roadmaps have their limitations in dynamic market environments. This is because roadmaps tend to shift the focus to features (e.g., prioritizing them for financial reasons), neglecting business goals or visions. Feature-based roadmaps are fixed-time-based charts that provide a forecast for specific products, features, or services (including concrete launch or deployment dates) [MTL19a]. Consequently, the focus often shifts to functions that do not contribute to the customer or business goals. Therefore, feature-based roadmaps are not suitable in dynamic market environments with the associated uncertainties but only in market environments that are predictable, stable, and reliable [MTL19a]. Hence, so-called dynamic roadmaps could be used. They describe how a company could achieve its vision and strategy to fulfill customer needs, connecting vision and needs with pathways to execute [MTL19b], [Tr21]. Roadmaps visualize possible routes to achieve a firms' vision and can be used as an artifact to raise awareness of internal and external stakeholders about the decisions made during the development [Lo17]. It should be noted that such roadmaps are appropriate to create alignment between the various actors in the ecosystem. In addition, roadmaps should be treated as revisable artifacts to cope with the changing conditions in dynamic environments. Depending on the purpose of the roadmap, there is a discussion on how to make the use of roadmaps more flexible, and there is already research on maturity models so that companies can evaluate whether they use roadmaps systematically and iteratively [MTL19b].

Nevertheless, the current scientific discourse on the use of roadmaps reveals that, although they are quite versatile, many companies still use them for long-range planning, which is rather static, and encounter difficulties in applying roadmaps according to the requirements of an uncertain and dynamic environment [MTL18], [MTL19b]. The dynamic roadmaps are precisely the ones that could potentially take on the ecosystem dynamics. However, few research studies have been conducted on the design of dynamic roadmaps, and they still relate to the individual organization level. Most roadmapping papers in the context of software describe the use of roadmaps for coordination within

individual organizations. However, roadmaps in dynamic networks such as ecosystems have hardly been researched. The missing application of roadmaps between the markets and single organization levels inspired us to highlight the fit between the characteristics and objectives of roadmaps and the challenges of ecosystem management, proposing the roadmapping approach for its ecosystem-wide use.

4 A Proposal for Using Roadmaps to Support Ecosystem Navigation

Ecosystem management is complex, cannot always be planned or remain directly controllable, and can eventually lead to undesired effects, whereby the absence of ecosystem management also increases the risk of ecosystem demise. This can be caused, for instance, by ruinous competition, which negatively impacts incentives for innovation, or by an increase in complementary products of insufficient quality [Bo12], [WFC14]. Both effects endanger the economic sustainability of an ecosystem. Thus, dynamic roadmaps deployed ecosystem-wide, offer the potential to promote alignment among multilateral ecosystem actors, achieving greater reliability in development and increasing ecosystem stability. Fig. 1. illustrates a proposal for using multi-layered ecosystem roadmaps to enable ecosystem actors to (1) start reflecting on the ecosystem-related vision, (2) then assessing the current state, and (3) the planning on how to achieve the vision.

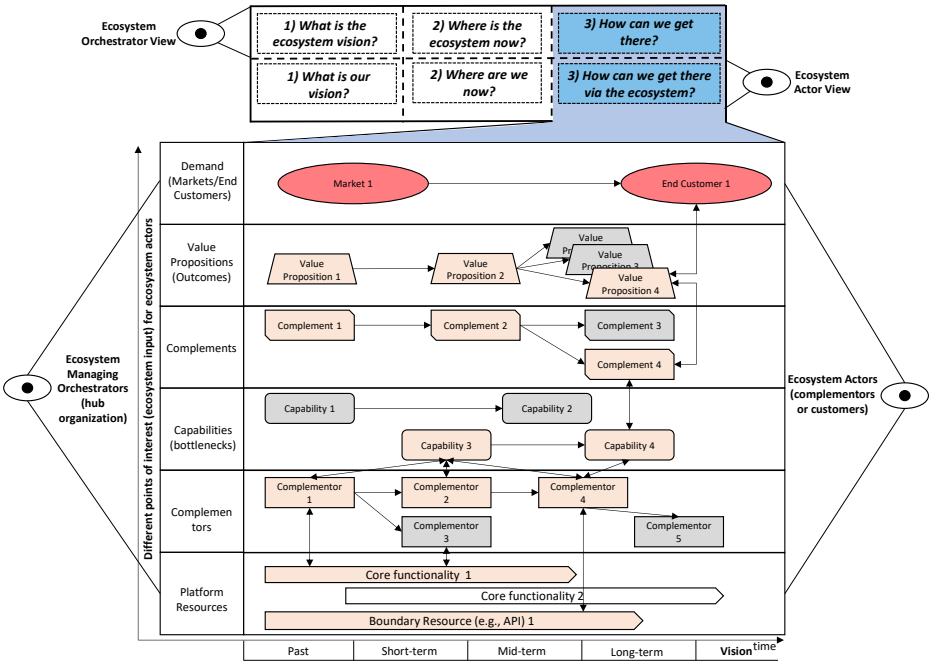


Fig. 1: Approach for ecosystem-wide use of roadmaps

As explained in the section before, different types of roadmaps can constitute a living artifact, which illustrates the causal relationships between different determinants that constitute ecosystems. That said, diverse ecosystem actors can see the critical pathways to position themselves to discover niches and recognize their own power position in the ecosystem. On the other hand, ecosystem orchestrators can use roadmaps to identify critical ecosystem actors, which could help to discover more ecosystem bottlenecks. Besides, ecosystem orchestrators can use roadmaps to improve value co-creation as the visualized added-value flows support matchmaking activities. Additionally, ecosystem orchestrators can use roadmaps to improve ecosystem communication, which is critical to defending significant governance decisions. If roadmaps are obligatory in ecosystems, power asymmetry cannot be prevented, but at least information asymmetry can be prevented. Imagine every ecosystem actor is required to document in the roadmap his decisions to abandon certain technologies, complements, or projects in the ecosystem. This would help increase transparency and trust about the competitive behavior of ecosystem actors among other ecosystem actors who rely on this abandoned technology. This would reduce the power asymmetry risks of both technological dependencies on platform providers (e.g., iOS and Flash deprecation) and complementary value-added partners (e.g., a unique complement). The beneficial use of roadmaps in ecosystems is based on the assumption that ecosystems cannot be designed but can be planned to a limited extent from the orchestrator's perspective. Planning can also help other ecosystem actors navigate (e.g., positioning) ecosystems. We suggest reading the roadmap from bottom to top in the ecosystem context. It starts with the illustration of platform boundary resources required by complementary ecosystem actors, which may contribute valuable capabilities to the ecosystem and create complements with a certain value proposition to fulfill the needs of end customers, representing the demanding side of the ecosystem [He20]. The time axis also allows for a specific mapping of ecosystem dynamics, which is often lacking in existing ecosystem management techniques [Ma21]. As a central artifact, ecosystem roadmaps can help visualize the ecosystem determinants relevant for joint value creation, and foster the discovery of critical bottlenecks or gaps. While ecosystem orchestrators could safeguard the bottlenecks, complementors could occupy the gaps in the offering. However, in contrast to earlier research, all ecosystem actors should view and collaboratively shape the roadmap across companies. In particular, using the roadmap as a living document for communicating strategic changes by individual ecosystem actors could prevent information and power asymmetries and increase transparency and trust. Based on roadmaps, the risks of potential and already participating ecosystem actors can be reduced through roadmaps, and companies' willingness to participate in the ecosystem can be promoted. Tab. 1 sums up the potential of roadmaps for ecosystem management.

After all, if roadmaps help to avoid certain risks for ecosystem actors, for example, increasing trust between organizations or even domains, research on ecosystem-wide use of roadmaps has the potential to foster the sustainability of ecosystems [Ga21], ensuring that complementors and customers benefit from mutual economic stability.

Ecosystem Challenges		Roadmap Potentials
Ecosystem structure	• Structure of interdependent actors • Formal independency • Oversupply • Diversity of actors • Power asymmetries • Information asymmetries	• Fostering the ecosystem transparency • Ecosystem actor alignment • Identification of product risks • Transparency on actors and capabilities • Visibility of dependencies • Increasing inter-organizational trust
Ecosystem management trade-offs	• Definition of decision rights • Input control • Provision of boundary resources • Preference for specific complements • Matchmaking	• Transparency on responsibilities • Outcome-based ecosystem control • Boundary resources management • Supporting discussion of preference decisions • Fostering actor assessment
Ecosystem dynamics	• Role instability • Bottlenecks • Sub ecosystems	• Ecosystem actor alignment • Communication of threats • Communication of threats
Ecosystem strategies	• Envelopment • Platform injection	• Transparency • Countering unwanted evolutions

Tab. 1: Potentials of roadmapping for ecosystem management

5 Future Steps

Some ecosystems also lack hub organizations, as the power differentials between organizations are small [TA18]. [DAA18], [HMP22]. Without hub organizations, ecosystems are considered to have an even higher degree of uncertainty for potential actors. Especially in such domains, the alignment of ecosystem actors with individual goals and agendas is problematic. We are confident that structuring ecosystem activities can also help the ecosystem firms better navigate without being in an orchestrating position. Ecosystem management is considered a critical capability for ecosystem navigation, and it needs to be acquired, especially by companies that were used to linear value chains and currently adapt to ecosystems. Roadmaps can, for instance, help sensitize industrial incumbents in contact with ecosystems. However, ecosystem roadmaps must be tailored for inter-organizational use and specific ecosystem determinants such as complements, ecosystem rules, or critical resources. Although tools for ecosystem management seem to exist already, most of these tools are not adapted to ecosystems

[Ma21]. In line with Matzner et al., we see the need for further research to test existing tools such as roadmaps for their suitability in an ecosystem-wide approach. Therefore, requirement elicitation studies with representatives from companies that create value in ecosystems is a promising future research direction. Empirical studies on the use of roadmaps in ecosystems could also be instantiated for the industrial domain, as prior research indicates several examples of ecosystem failure there. Considering industrial firms' rigorous requirements to participate in ecosystems [PFM21], empirical validation of the roadmap suitability in general and its adoption for ecosystem navigation in the industrial domain would be valid for future research, as there is currently no application example for dynamic roadmaps in ecosystems. In addition, an analysis of software ecosystems with release roadmaps for developers (e.g., Apple iOS) could be helpful in determining the status quo in ecosystem communication and use it as a basis for future design-oriented research on ecosystem roadmaps.

Bibliography

[Ad17] Adner, R.: Ecosystem as Structure: An Actionable Construct for Strategy. Journal of Management, 43/1, pp. 39-58, 2017.

[AK03] Albright, R.E., Kappel, T.A.: Roadmapping In the Corporation. Research-Technology Management 46/2, pp. 31-40, 2003.

[Au21] Autio, E. Orchestrating ecosystems: a multi-layered framework. Innovation, 1-14, 2021.

[Bo12] Boudreau, K.J.: Let a thousand flowers bloom? An early look at large numbers of software app developers and patterns of innovation. Organization Science, 23/5, pp. 1409-1427, 2012.

[BO18] Bosch, J. and Olsson, H.H.: Ecosystem traps and where to find them. J. Softw. Evol. Proc. 30/11, 2018.

[CF21] Cenamor, J., Frishammar, J.: Openness in platform ecosystems: Innovation strategies for complementary products. Research Policy 50, 104148, 2021.

[CGY19] Cusumano, M., Gawer, A., Yoffie, D.: The Business of Platforms: Strategy in the Age of Digital Competition, Innovation, and Power. Harper Business, 2019.

[CK21] Cutolo, D., Kenney, M.: Platform-dependent entrepreneurs: Power asymmetries, risks, and strategies in the platform economy. Academy of Management Perspectives 35/4, pp. 584-605, 2021.

[DAA18] Dattée, B., Alexy, O, Autio, E.: Maneuvering in poor visibility: how firms play the ecosystem game when uncertainty is high. Academy of Mangement Journal 61/2, pp. 466-498, 2018.

[DP06] Dhanaraj, C., Parkhe, A.: Orchestrating innovation networks. Academy of management review, 31/3), pp. 659-669, 2006.

[Ea15] Eaton, B. D. et.al.: Distributed tuning of boundary resources: the case of Apple's iOS service system. MIS Quarterly, 39/1, pp. 217–243, 2015.

[Ga14] Gawer, A.: Bridging differing perspectives on technological platforms: Toward an integrative framework. Research policy 43/7, pp. 1239-1249, 2014.

[Ga21] Gawer, A.: Digital platforms and ecosystems: remarks on the dominant organizational forms of the digital age. Innovation: Org. & Man. 24/1, pp. 110-124, 2021.

[GC13] Gawer, A., Cusumano, M.A.: Industry Platforms and Ecosystem Innovation. J. Prod. Innov. Manag. 31/3, pp. 417-433, 2013.

[Go21] Gomes, L.A.d.V., Flechas, X.A., Facin, A.L.F., Borini, F.M.: Ecosystem management: Past achievements and future promises. Technol. Forecasting & Soc. Change 171, 120950, 2021.

[Gu20] Guggenberger, T.M., Möller, F., Haarhaus, T., Gür, I., Otto, B.: Ecosystem Types in Information Systems. In: Proceedings of the 28th ECIS, Marrakesh, 2020.

[He20] Hein, A. et.al.: Digital platform ecosystems. Electr. Mark. 30, pp. 87-98, 2020.

[HE17] Hannah, D.P., Eisenhardt, K.M.: How firms navigate cooperation and competition in nascent ecosystems. Strategic Management Journal, 39/12, pp. 3163-3192, 2017.

[HHD21] Hurni, T., Huber, T.L., Dibbern, J.: Power dynamics in software platform ecosystems. Information Systems Journal, 32/2, pp. 310-343, 2021.

[HMP22] Hiller, S., Morar, D., Petrik, D.: Creating Value in Additive Manufacturing – Modeling of Ecosystem Determinants. In: Proceedings of the 30th ECIS, Timisoara, 2022.

[HW22] Heimburg, V., Wiesche, M.: Reklations between Actors in Digital Platform Ecosystems: A Literature Review. In: Proceedings of the 30th ECIS, Timisoara.

[IL04] Iansiti, M., Levien, R.: The keystone advantage: What the new dynamics of business ecosystems mean for strategy, innovation, and sustainability. Boston Mass., Harvard Business School Press, pp. 1-10, 2004.

[JCG18] Jacobides, M.G., Cennamo, C., Gawer, A.: Towards a theory of ecosystems. Strat. Man. J. 39/8, pp. 2255-2276, 2018.

[JTS20] Jablonski, M., Timmers, P., Sarkis, J.: Sustainability in business models in the network economy. Electr. Mark. 30, pp. 675-678, 2020.

[JSP21] Jovanovic, M., Sjödin, D., Parida, V.: Co-evolution of platform architecture, platform services, and platform governance: Expanding the platform value of industrial digital platforms. Technovation, 102218, 2021.

[Ka01] Kappel, T.A.: Perspecives on roadmaps: how organizations talk about the future. The Journal of Product Innovation Management 18, pp. 39-50, 2001.

[KBZ21] Kenney, M., Bearson, D., Zysman, J.: The platform economy matures: measuring pervasiveness and exploring power. Socio-Economic Review, 00, pp. 1-33, 2021.

[KR20] Karhu, K., Rittala, P.: Slicing the cake without baking it: Opportunistic platform entry strategies in digital markets. L. R. Plan, 101988, 2020.

[LN15] Lusch, R.F., Nambisan, S: 2015: Service Innovation: A Service-Dominant Logic Perspective.MIS Quarterly 39/1, pp. 155-175, 2015.

[Lo17] Lombardo, C. T. et.al.: Product roadmaps relaunched - How to set direction while embracing uncertainty, O'Reilly Media, Sebastopol, 2017.

[LKK05] Lehtola, L.; Kauppinen, M.; Kujala, S.: Linking the business view to requirements engineering: Long-term product planning by roadmapping. In (IEEE): Proceedings of the 13th IEEE International Conference on Requirements. Paris, pp. 439-443, 2005.

[Ma21] Matzner, M., Pauli, T., Marx, E., Anke, J., Poeppelbuss, J., Fielt, E., Gregor, S., Sun, R., Hydle, K.M., Aas, T.H., Aanestad, M., Gordijn, J., Kaya, F., Wieringa, R.: Transitioning to Platform-based Services and Business Models in a B2B Environment. Journal of Service Management Research 5/2, pp. 143-162, 2021.

[MI17] Möhrle, M.G., Isenmann, R.: Grundlagen des Technologie-Roadmapping. In (Möhrle, M.G., Isenmann, R. Eds.) Technologie-Roadmapping. Berling, pp. 1-17, 2017.

[MTL18] Münch, J.; Trieflinger, S.; Lang, D.: Why feature-based roadmaps fail in rapidly changing markets - A qualitative suvey. In (Hyrynsalmi, S., Ed.): Proc. Int. Workshop on Software-intensive Business - Start-ups, Ecosystems and Platforms. Aachen, pp. 202-218, 2018.

[MTL19a] Münch, J.; Trieflinger, S.; Lang, D.: What's hot in product roadmapping? Key practices and success factors. In (Franch, S., Ed) Proc. 20th International Conference on Product-Focused Software Process Improvement. Cham, pp. 401-416, 2019.

[MTL19b] Münch, J.; Trieflinger, S.; Lang, D.: The product roadmap maturity model DEEP: Validation of a method for assessing the product roadmap capabilities of organizations. In (Hyrynsalmi, S., ed.): Proc. 10th International Conference on Software Business. Cham, pp. 97-113, 2019.

[Mo93] Moore, J.F.: Predators and prey: a new ecology of competition. Harvard Bus. Rev. 71/3, S. 75-86, 1993.

[Mo96] Moore, J.F.: The death of competition: leadership and strategy in the age of business ecosystems. HarperCollins, 1996.

[OK20] O'Mahony, S., Karp, R.: From proprietary to collective governance: How do platform participation strategies evolve?. Strategic Management Journal 43/3, pp. 530-562, 2020.

[PAJ17] Parker, G., Van Alstyne, M., W., Jiang X.: Platform Ecosystems: How Developers Invert the Firm. MIS Quarterly 41/1, pp. 255-266, 2017.

[PFM21] Pauli, T., Fielt, E., Matzner, M.: Digital Industrial Platforms. Bus Inf Syst Eng 63/2, pp. 181-190, 2021.

[PFP01] Phaal, R., Farrukh, C., Probert, D.: Technology Roadmapping: linking technology resources to business objectives. Centre for Technology Management, University of Cambridge, pp. 1-18, 2001.

[PFP04] Phaal, R., Farrukh, C., Probert, D.: Customizing Roadmapping. Research-Technology Management 47/2, pp. 26-37, 2004.

[PH14] Porter, M.E., Heppelmann, J.E.: How smart connected products are transforming competition. Harvard Bus. Rev. 92/11, pp. 64–88, 2014.

[PS21] Petrik, D., Schüler, F.: Einfluss plattformbasierter Ökosysteme auf unternehmensübergreifende integrierte Wertschöpfungsnetzwerke. In (Schulz, T. eds.): Wertschöpfungsnetzwerke mit digitalisierten Dienstleistungen etablieren. Beuth Verlag, Berlin, pp. 85-103, 2021.

[RS21] Rietveld, J., Schilling, M.A.: Platform competition: A systematic and interdisciplinary review of the literature. Journal of Management, 47/6, pp. 1528-1563, 2021.

[SG20] Shipilov, A., Gawer, A.: Integrating research on interorganizational networks and ecosystems. Academy of Management Annals, 14/1, pp. 92-121, 2020.

[SHL20] Sandberg, J., Holmström, J., Lyytinen, K.: Digitization and Phase Transitions in Platform Organizing Logics: Evidence from the Process Automation Industry. MIS Quarterly. 44/1, pp. 129-153, 2020.

[Sk19] Sklyar, A., Kowalkowski, C., Tronvoll, B, Sörhammar, D.: Organizing for digital servitization: A service ecosystem perspective. Journal of Bus. Res. 104, pp. 450-460, 2019.

[SP21] Schüler, F., Petrik, D.: Objectives of Platform Research: A Co-Citation and Systematic Literature Review Analysis. In (Seiter, M., Grünert, L., Steur, A., eds.): Management Digitaler Plattformen. Springer Gabler, Wiesbaden, pp. 1-33 2021.

[SKD18] Stummer, C., Kundisch, D., Decker, R.: Platform Launch Strategies. Bus Inf Syst Eng 60/2, pp. 167-173, 2018.

[Su11] Suomalainen, T., Salo, O., Abrahamsson, P., Similä, J.: Software product roadmapping in a volatile business environment. He Journal of Systems and Software 84, pp. 958-975, 2011.

[SWK16] Schreieck, M., Wiesche, M., Krcmar, H.: Design and Governance of Platform Ecosystems – Key Concepts and Issues for Future Researc. Proceedings of the 24th ECIS, Istanbul.

[TA18] Thomas, L.D.W., Autio, E.: Ecosystem Value Potential: An Organizational Field Perspective. Academy of Management Annual Meeting Proceedings, 2018.

[Ti14] Tiwana, A.: Platform ecosystems: Aligning architecture, governance, and strategy. Morgan Kaufmann, 2014.

[TKB10] Tiwana, A., Konsynski, B., Bush, A.A.: Platform Evolution: Coevolution of Platform Architecture, Governance, and Environmental Dynamics. Inf. Syst. Res. 21/4, pp. 675-687, 2010.

[Tr21] Trieflinger, S., Münch, J., Wagner, S., Lang, D., Roling, B.: A Transformation Model for Excelling in Product Roadmapping in Dynamic and Uncertain Market Environments. In (Ardito, L., Jedlitschka, A., Morisio, M., Torchiano, M. eds.): Product-Focused Software Process Improvement. PROFES 2021, Cham. pp. 136-151, Springer, 2021.

[Va15] Valkokari, K.: Business, Innovation, and Knowledge Ecosystems: How They Differ and How to Survive and Thrive within Them. Technology Innovation Management Review, 5/8, pp. 17-24, 2015.

[Vi21] Vinayavekhin, S., Phaal, R., Thanamaitreejit, T., Asatani, K.: Emerging Trends in Roadmapping Research: A Bibliometric Literature Review. Technology Analysis & Strategic Management, 2021.

[VLR02] Vähäniitty, V.; Lassenius, C.; Rautiainen, K.: An approach to product roadmapping in small software product businesses. In (Kontio, J., Conradi, R., eds.): Proc. 7th Int. Conf. on Software Quality. Helsinki, pp. 12-13, 2002.

[WFC14] Wareham, J., Fox, P. B., Cano Giner, J. L.: Technology ecosystem governance. Organization science, 25/4, pp. 1195-1215, 2014.

[WN13] Weiller, C. Neely, A.: Business model design in an ecosystem context. University of Cambridge, Cambridge Service Alliance, pp. 1–21, 2013.

[WW14] West, J., Wood, D.: Evolving an open ecosystem: The rise and fall of the Symbian platform. In (Adner, R., Oxley, J.E., Silverman, B.S. ed.): Collaboration and competition in business ecosystems, pp. 27-67, Emerald Group Publishing Limited, 2014.

Konzeptionelle Herausforderungen für die KI: Naturwissenschaften

Konzeptionelle Grundlagen der Künstlichen Intelligenz: KI und Naturwissenschaften

Klaus Mainzer[1,2]

Abstract: Im ersten Lehrbuch der neuzeitlichen Physik „Mathematische Prinzipien der Naturphilosophie“ (Principia Mathematica Philosophiae Naturalis) beschreibt Newton die Methoden der Naturforschung (regulae philosophandi), die bis heute die Forschungspraxis bestimmen: Die Naturforschung soll zunächst „induktiv“ aus Beobachtungs- und Messdaten kausale Zusammenhänge ableiten und in mathematischen Gesetzen formulieren, aus denen dann „deduktiv“ Erklärungen und Prognosen abgeleitet werden (1). Die deduktive Methode wurde Anfang des 20. Jahrhunderts in der mathematischen Logik formalisiert und Grundlage in der ersten KI-Phase des automatischen Beweisens und der Expertensysteme („symbolic AI“) (2). Die induktive Methode wird heute in der KI auf der Grundlage statistischer Lerntheorie zur Entdeckung von Datenkorrelationen verwendet und ist Grundlage des modernen Machine Learning. Statistische Korrelationen ersetzen aber keine kausalen Beziehungen (3). Damit hängen praktische Sicherheits- und Verifikationsfragen von Computerprogrammen eng zusammen. In den modernen Naturwissenschaften von der Physik und Chemie bis zur Biologie und den Life Sciences spielen statistische Methoden eine zentrale Rolle.

KI unterstützt aber nicht nur Forschungsmethoden in den Naturwissenschaften. Umgekehrt werden heute Methoden der mathematischen Physik verwendet, um die gewaltigen neuronale Netze der KI (Deeper Learning) in Wissenschaft und technischer Praxis zu verstehen und rechnerisch zu beherrschen (4). KI und Naturwissenschaften wachsen also in jeder Hinsicht methodisch zusammen.

Abstract (English): In the first textbook of modern physics, "Mathematical Principles of Natural Philosophy" (Principia Mathematica Philosophiae Naturalis), Newton describes the methods of natural research (regulae philosophandi) that still determine research practice today: Natural research should first "inductively" derive causal relationships from observation and measurement data and formulate them in mathematical laws, from which explanations and predictions are then "deductively" derived (1). The deductive method was formalised in mathematical logic at the beginning of the 20th century and became the basis in the first AI phase of automatic reasoning and expert systems ("symbolic AI") (2). Today, the inductive method is used in AI based on statistical learning theory to discover data correlations and is the basis of modern machine learning. However, statistical correlations do not replace causal relationships (3). Thus, practical security and verification issues of computer programmes are closely related. Statistical methods play a central role in the modern natural sciences, from physics and chemistry to biology and the life sciences.

However, AI is not only applied in the natural sciences. Conversely, methods of mathematical physics are now used to understand and computationally master the vast neural networks (Deeper Learning) in science and technical practice (4). AI and the natural sciences are thus methodically growing together in every respect.

[1] TUM Senior Excellence Faculty, Technische Universität München
[2] Carl Friedrich von Weizsäcker Center, Universität Tübingen

doi:10.18420/inf2022_36

1. Newtons Methodenlehre der Naturwissenschaften

Newtons Methodenlehre der Naturforschung, die in seinem Lehrbuch „Principia Mathematica Philosophiae Naturalis“ als „regulae philosophandi“ (Regeln des Philosophierens) bezeichnet wird, ist einfach und hat die nachfolgenden Jahrhunderte der Physik geprägt. Sie wird nahezu wörtlich von Einstein u.a. im 20. Jahrhundert wiederholt: Der Naturforscher beginnt mit Beobachtungen und Messungen und erkennt in diesen endlich vielen Daten Zusammenhänge, die er in der Annahme eines Gesetzes oder einer Theorie aus mehreren Gesetzen verallgemeinert. Mathematisch haben sie häufig die Form von Gleichungen wie z.B. die Bewegungs- und Kraftgleichungen in der Newtonschen Mechanik. In der Wissenschaftstheorie nennen wir solche Gesetzesannahmen auch „Hypothesen“. Eine empirische Theorie ist danach ein System von Gesetzeshypothesen. Die Gesetzeshypothesen entsprechen den Axiomen in einer mathematischen Theorie. Den Weg von den empirischen Daten zu Hypothesen oder Theorien nannte Newton „Induktion“.[3] In der (modernen) Mathematik sind Axiome nur Setzungen und Annahmen, aus denen Theoreme logisch abgleitet werden. Ob und wie sie sich aus empirischen Daten begründen lassen, ist nicht Gegenstand der Logik und Mathematik.

Aus empirischen Theorien und Hypothesen lassen sich für geeignete Anfangs- und Nebenbedingungen Voraussagen für zukünftige Ereignisse oder Erklärungen für vergangene und gegenwärtig beobachtete Ereignisse logisch-mathematisch ableiten. Ein Beispiel ist Keplers Planetenmodell, das aus Newtons Mechanikgesetzen ableitbar ist. Unter seiner Annahme lassen sich Prognosen über zukünftige Planetenstellungen ableiten. Wir setzen voraus, dass eine Aussage *A* (z.B. Keplers Planetengesetze mit einer bekannten Anfangsposition eines Planeten) eine Aussage *B* (z.B. eine Prognose über eine zukünftige Planetenposition) logisch impliziert. Die Wahrheit dieser Implikation wird durch Ausrechnung der Bewegungsgleichung des Planeten nach Einsetzung seiner Anfangsposition bewiesen. Nun wird die Aussage *A*, also Keplers Planetenmodell und die Anfangsposition des Planeten, als wahr vorausgesetzt. Dann ist logisch zwingend auch Aussage *B*, also die Prognose über den zukünftigen Planetenort wahr: Wenn der Schluss „Wenn *A*, dann *B*“ wahr und die Aussage *A* wahr sind, dann ist logisch zwingend die Aussage *B* wahr.

Nun nehmen wir aber an, dass der Schluss „Wenn *A*, dann *B*“ wahr und die Aussage *B* (z.B. eine Beobachtung des Planetenorts) wahr seien. Was wissen wir dann über das angenommene Gesetz bzw. Modell? Jedenfalls ist Aussage *A* logisch nicht zwingend wahr. Nach gehäuften Beobachtungen, wonach die logischen Folgerungen *B* aus *A* wahr sind, könnte man bestenfalls vermuten, dass unser angenommenes Modell bzw. die Hypothese *A* plausibel ist. Offensichtlich ist das die Situation in den empirischen Wissenschaften wie Physik, Chemie und Biologie, aber auch Sozial- und Wirtschaftswissenschaften.

[3] J. Mittelstrass (1995), Regulae philosophandi, in: J. Mittelstrass (Hrsg.), Enzyklopädie Philosophie und Wissenschaftstheorie Bd. 3, J.B. Metzler: Stuttgart, 536-537.

2. Logische Deduktion und symbolische KI in den Naturwissenschaften

2.1 Symbolische KI und automatisches Beweisen

In der ersten Phase der KI-Forschung war die Suche nach allgemeinen logischen Problemlösungsverfahren wenigstens in der formalen Logik erfolgreich. Dort wurde ein mechanisches Deduktionsverfahren angegeben, um die logische Allgemeingültigkeit von Formeln zu beweisen. Das Verfahren konnte auch von einem Computerprogramm ausgeführt werden und leitete in der Informatik das automatische Beweisen ein.[4]

Der Grundgedanke ist einfach zu verstehen. In der Algebra werden Buchstaben x, y, z, … durch Rechenoperationen wie z. B. Addieren (+) oder Subtrahieren (–) verbunden. Die Buchstaben dienen als Leerstellen (Variablen), um Zahlen einzusetzen. In der formalen Logik werden Aussagen durch Variablen A, B, C, … bezeichnet, die durch logische Junktoren wie z. B. ‚und' (∧), ‚oder' (∨), wenn-dann' (→), ‚nicht' (¬) verbunden werden. Die Aussagenvariablen dienen als Leerstellen, um Aussagen einzusetzen, die entweder wahr oder falsch sind. So ist z. B. $A \wedge B$ eine logische Formel, die durch Einsetzung der wahren Aussagen 1+3=4 für A und 4=2+2 für B in die wahre Aussage $1+3=4 \wedge 4=2+2$ übergeht. In der Arithmetik ergibt sich daraus der wahre Schluss $1+3=4 \wedge 4=2+2 \rightarrow 1+3=2+2$. Allgemein ist aber der Schluss $A \wedge B \rightarrow C$ nicht wahr. Demgegenüber ist der Schluss $A \wedge B \rightarrow A$ logisch allgemeingültig, da für die Einsetzung von beliebigen wahren oder falschen Aussagen für A und B sich immer eine wahre Gesamtaussage ergibt.

Der Beweis für die Allgemeingültigkeit eines logischen Schlusses kann in der Praxis sehr kompliziert sein. Daher schlug John A. Robinson 1965 die sogenannte Resolutionsmethode vor, nach der Beweise nach dem Muster logischer Widerlegungsverfahren gefunden werden können.[5] Man startet also mit der Annahme des Gegenteils (Negation), d. h. der logische Schluss sei nicht allgemeingültig. Im nächsten Schritt wird dann gezeigt, dass alle möglichen Anwendungsbeispiele dieser Annahme zu einem Widerspruch führen. Es gilt also das Gegenteil der Negation und der logische Schluss ist allgemeingültig. Die Robinsonsche Resolutionsmethode benutzt logische Vereinfachungen, wonach man jede logische Formel in eine sogenannte konjunktive Normalform umwandeln kann. In der Aussagenlogik besteht eine konjunktive Normalform aus negierten und nicht negierten Aussagenvariablen (Literale), die durch Konjunktion (∧) und Disjunktion (∨) verbunden sind.

Mechanisch besteht das Verfahren also darin, widersprechende Teilaussagen aus Konjunktionsgliedern einer logischen Formel zu streichen („Resolution") und diese Prozedur mit den entstehenden „Resolventen" und einem entsprechenden anderen Konjunktionsglied der Formel zu wiederholen, bis ein Widerspruch (die „leere" Klausel, die nie wahr werden kann) abgeleitet werden kann.

In einem entsprechenden Computerprogramm terminiert dieses Verfahren für die Aussagenlogik. Es zeigt also in endlicher Zeit, ob die vorgelegte logische Formel

[4] Mainzer (2019) (Ref.), Kap. 3.
[5] J.A. Robinson (1965), A machine oriented logic based on the resolution principle, in: *Journal of the Association for Computing Machinery* 12:23–41.

allgemeingültig ist. Allerdings wächst die Rechenzeit nach den bisher bekannten Verfahren exponentiell mit der Anzahl der Literale einer Formel. Was nun die „Künstliche Intelligenz“ betrifft, so können Computerprogramme mit der Resolutionsmethode automatisch über die Allgemeingültigkeit logischer Schlüsse wenigstens in der Aussagenlogik im Prinzip entscheiden. Menschen hätten große Schwierigkeiten, bei komplizierten und langen Schlüssen den Überblick zu bewahren. Zudem sind Menschen wesentlich langsamer. Mit steigender Rechenkapazität können Maschinen also wesentlich effizienter diese Aufgabe des logischen Schließens erledigen.

In den Naturwissenschaften haben wir es aber mit komplexen prädikativen Aussagen zu tun. In der Prädikatenlogik werden Aussagen in Eigenschaften (Prädikate) zerlegt, die Objekten zu- oder abgesprochen werden. So wird in der Aussage $P(a)$ (z. B. ‚Atomkern a ist schwer‘) einem Atomkern namens a das Prädikat ‚schwer‘ (P) zugesprochen. Diese Aussage ist wieder entweder wahr oder falsch. In einer prädikativen Aussageform $P(x)$ werden Leerstellen (Individuenvariablen) x, y, z, … verwendet, für die Individuen a, b, c, … einer angenommenen Anwendungsdomäne (z. B. Atomkerne eines Reaktors) eingesetzt werden können. Neben den logischen Verknüpfungen der Aussagenlogik können nun auch Allquantoren $\wedge x$ (‚Für alle x‘) und Existenzquantoren $\vee x$ (‚Es gibt ein x‘) verwendet werden.

Für die Formeln der Prädikatenlogik lässt sich ebenfalls ein verallgemeinertes Resolutionsverfahren angeben, um wieder aus der Annahme der allgemeinen Ungültigkeit einer Formel einen Widerspruch abzuleiten. Dazu muss eine Formel der Prädikatenlogik ähnlich wie in der Aussagenlogik (aber deutlich komplizierter) in eine Normalform gebracht werden, aus deren Klauseln sich mechanisch ein Widerspruch ableitet lässt. Falls die Formel tatsächlich widersprüchlich ist, kommt das Verfahren auch in endlicher Zeit zu diesem Ergebnis. Allgemeingültigkeit kann man gleichermaßen in endlicher Zeit nachweisen, da in diesem Fall die Negation der Formel widersprüchlich ist (es gibt kein Gegenbeispiel). Allerdings gibt es einen dritten Fall in dem die Formel in manchen Anwendungsfällen wahr ist und in anderen falsch, je nachdem wie man die Individuen und Prädikate interpretiert. Daher kann es in der Prädikatenlogik (im Unterschied zur Aussagenlogik) vorkommen, dass das Resolutionsverfahren kein Ende findet. Das Computerprogramm läuft dann unbegrenzt weiter. Es kommt dann darauf an, Teilklassen zu finden, in denen das Verfahren nicht nur effizient läuft, sondern überhaupt terminiert.

Die Schule um den Mathematiker David Hilbert hatte bereits in den 1920er-Jahren angenommen, dass Prädikatenlogik die „Sprache der Mathematik“ sei. Danach müssten auch mathematische Anwendungsgebiete wie z.B. die Physik im Prinzip mit Prädikatenlogik formuliert werden können. Gleichzeitig hatte die mathematische Logik ein (zunächst sehr theoretisches) Verfahren entwickelt, um die Allgemeingültigkeit (bzw. Widersprüchlichkeit) einer Formel mechanisch nachzuweisen. Eine Formel kann nur dann widersprüchlich sein, wenn es ein endliches Anwendungsbeispiel gibt, das den Widerspruch aufzeigt. Logisch geht es also um die Erfüllbarkeit (satisfiability) oder Widersprüchlichkeit von Formeln, die durch mechanische Algorithmen (SAT-Algorithmen) zu entscheiden sind.[6] Die

[6] A. Biere, M. Heule, H. van Maaren, T. Walsh (Hrsg.) (2009), Handbook of satisfiability. IOS Press: Amsterdam.

KI, die sich am Paradigma der symbolischen Logik orientiert, wird heute „symbolische KI“ genannt.

2.2 Symbolische KI und Expertensysteme in den Naturwissenschaften

Naturwissenschaftler*nnen sind Expert*nen, die über ein spezielles Domänenwissen der Naturwissenschaften verfügen. Wissensbasierte Expertensysteme sind Computerprogramme, die Wissen über ein spezielles Gebiet speichern und ansammeln, aus dem Wissen automatisch Schlussfolgerungen ziehen, um zu konkreten Problemen des Gebietes Lösungen anzubieten. Im Unterschied zum menschlichen Experten ist das Wissen eines Expertensystems aber auf eine spezialisierte Informationsbasis beschränkt ohne allgemeines und strukturelles Wissen über die Welt.[7]

Um ein Expertensystem zu bauen, muss das Wissen des Experten in Regeln gefasst werden, in eine Programmsprache übersetzt und mit einer Problemlösungsstrategie bearbeitet werden. Expertensysteme sind damit ein typisches Beispiel für symbolische KI.

Historisch war DENDRAL eines der ersten erfolgreichen Expertensysteme in den Naturwissenschaften, das E.A. Feigenbaum u.a. Ende der 60er Jahre in Stanford entwickelten.[8] Es benutzt die speziellen Kenntnisse eines Chemikers, um zu einer chemischen Summenformel eine passende molekulare Strukturformel zu finden. In einem ersten Schritt werden systematisch alle mathematisch möglichen räumlichen Anordnungen der Atome zu einer vorgegebenen Summenformel bestimmt. Für z.B. $C_{20}H_{43}N$ ergeben sich 43 Millionen Anordnungen. Chemisches Wissen über die Bindungstopologie, wonach z.B. Kohlenstoff-Atome vielfach gebunden werden können, reduzieren die Möglichkeiten auf 15 Millionen. Wissen über Massenspektroskopie, über die wahrscheinlichste Stabilität von Bindungen (heuristisches Wissen) und Kernspinresonanz schränken schließlich die Möglichkeiten auf die gesuchte Strukturformel ein.

Die Problemlösungsstrategie, die hier zugrunde gelegt wurde, ist offenbar nichts anderes als der vertraute „British-Museum-Algorithm“, der in der Programmiersprache LISP formuliert wurde. Das Verfahren lautet also GENERATE_AND_TEST, wobei im GENERATE-Teil die möglichen Strukturen systematisch erzeugt werden, während die chemische Topologie, Massenspektroskopie, chemische Heuristik und Kernspinresonanz jeweils Test-Prädikate angeben, um die möglichen Strukturformeln einzuschränken.

Zweckmäßigerweise lassen sich Problemlösungstypen in Diagnose-, Konstruktions- und Simulationsaufgaben einteilen. Typische diagnostische Problembereiche sind medizinische Diagnostik, technische Diagnostik wie z.B. Qualitätskontrolle, Reparaturdiagnostik oder Prozessüberwachung und Objekterkennung. Daher löst auch DENDRAL ein

[7] L.F. Puppe (1988), Einführung in Expertensysteme, Springer: Berlin.

[8] B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum (1969), Heuristic DENDRAL: A program for generating processes in organic chemistry, in: B. Meltzer, D. Michie (eds.), Machine Intelligence 4, Edinburgh; B. G. Buchanan, E.A. Feigenbaum (1978), DENDRAL and META-DENDRAL: Their applications dimensions, in: Artificial Intelligence 11, 5-24.

typisches Diagnoseproblem, indem es nämlich die passende molekulare Struktur für eine vorgegebene Summenformel erkennt.

Das erste medizinische Beispiel eines Expertensystems war MYCIN, das Mitte der 1970er Jahre an der Universität Standford entwickelt wurde.[9] Das MYCIN-Programm wurde zur medizinischen Diagnose geschrieben, um einen Arzt mit medizinischem Spezialwissen über bakterielle Infektion zu simulieren. Methodisch handelt es sich um ein Deduktionssystem mit Rückverkettung. MYCINs Wissenspool über bakterielle Infektionen besteht aus etwa 300 Produktionsregeln.

Um das Wissen anwenden zu können, arbeitet MYCIN rückwärts. Für jede von 100 möglichen Hypothesen von Diagnosen versucht MYCIN auf einfache Fakten zu stoßen, die durch Laborergebnisse oder Klinikbeobachtungen bestätigt sind. Da MYCIN in einem Bereich arbeitet, in dem Deduktionen kaum sicher sind, wurde eine Theorie des plausiblen Schließens und der Wahrscheinlichkeitsbewertung mit dem Deduktionsapparat verbunden. Es handelt sich dabei um sogenannte Sicherheitsfaktoren für jeden Schluss in einem AND/OR-Baum.

Dort bezeichnet F_i den Sicherheitsfaktor, den ein Benutzer einer Tatsache zumisst. C_i gibt den Sicherheitsfaktor eines Schlusses an, A_i den Grad der Verlässlichkeit, der einer Produktionsregel zugetraut wird. An den AND- bzw. OR-Knoten werden jeweils Sicherheitsfaktoren der entsprechenden Formel berechnet. Falls der Sicherheitsfaktor einer Datenangabe nicht größer als 0,2 sein sollte, so gilt sie als unbekannt und erhält den Wert 0. Das Programm berechnet also Bestätigungsgrade in Abhängigkeit von mehr oder weniger sicheren Fakten. MYCIN wurde unabhängig von seiner speziellen Datenbasis über Infektionskrankheiten für verschiedene diagnostische Anwendungsbereiche verallgemeinert. Wissenschaftstheoretisch waren Autoren von MYCIN wie B.G. Buchanan von R. Carnaps Logik der Induktion inspiriert.

Praktische Grenzen von Expertensystemen wurden schrittweise ausgeweitet: So zeichnen sich Experten nicht dadurch aus, dass sie mit absoluter Sicherheit zwischen wahr und falsch unterscheiden können und damit größere Genauigkeit vortäuschen, als erreichbar ist. Ein guter Experte vermag vielmehr, Unsicherheiten einzuschätzen, die sich z.B. bei der medizinischen Diagnose in der Symptomerhebung oder Symptombewertung ergeben. In Expertensystemen wird daher häufig nicht die klassische Logik mit der Annahme der Zweiwertigkeit der Wahrheitswerte („tertium non datur“) zugrunde gelegt, sondern zusätzlich Unsicherheitswerte wie z.B. „sicher“, „wahrscheinlich“, „möglich“ u.a. angenommen. Es ist ein altes Problem der Wissenschaftstheorie, dass nur logische Schlüsse mit Sicherheit gelten, also z.B. der direkte Schluss, der aus der Annahme von $A \rightarrow B$ und A die Wahrheit von B folgert. A ist in diesem Fall eine hinreichende Bedingung für B, während B für A nur notwendig ist. Daher kann aus $A \rightarrow B$ und B nicht logisch zwingend auf

[9] E.H. Shortliffe (1976), Computer-Based Medical Consultations: MYCIN, New York; D. Randall, B.G. Buchanan, E.H. Shortliffe (1977), Producing rules as a representation for a knowledge-based consultation program, in: Artificial Intelligence 8.

A geschlossen werden. Damit stellt sich in den Naturwissenschaften wieder das Newtonsche Induktionsproblem.

3. Empirische Induktion und subsymbolische KI

Im Unterschied zu Logik und Mathematik stehen im Zentrum der empirischen Wissenschaften die Fragen: Wie findet man in Datenmengen geeignete Modelle für Erklärungen und Prognosen? Wie plausibel sind solche Modelle? Wie ändert sich ihre Glaubwürdigkeit bei neuen Beobachtungen und Hintergrundwissen? Was können wir aus diesen neuen Erfahrungen lernen? Wissenschaftstheoretisch geht es also um Newtons altes Induktionsproblem. Mit den ungeheuren Datenmengen, mit denen es z.B. Teilchenphysik und Molekularbiologie zu tun haben, wird klar, dass diese Fragen nur im mathematischen Rahmen von Statistik und Wahrscheinlichkeitstheorie geklärt werden können.

3.1 Empirische Induktion in der statistischen Lerntheorie

Die Wahrscheinlichkeitstheorie geht vom Modell eines Zufallsexperiments aus. Mathematisch wird dazu ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ angenommen, wobei Ω die Menge aller Ergebnisse des Zufallsexperiments, $\mathcal{F}$ die Menge der Ereignisse $A \subseteq \Omega$ und P das Maß ist, das jedem Ereignis eine Wahrscheinlichkeit zuordnet. Probabilistisches Schließen versucht, Eigenschaften der Resultate von Zufallsexperimenten aus der gegebenen mathematischen Struktur $(\Omega, \mathcal{F}, P)$ abzuleiten. Probabilistisches Lernen versucht, Eigenschaften des zugrunde gelegten statistischen Modells aus den Resultaten (Daten) des Experiments abzuleiten.

Beim probabilistischen bzw. statistischen Lernen sollen also Abhängigkeiten und Zusammenhänge aus Beobachtungsdaten durch Algorithmen abgeleitet werden. Dazu können wir uns ein naturwissenschaftliches Experiment vorstellen, bei dem in einer Serie von veränderten Versuchsbedingungen (Inputs) entsprechende Ergebnisse (Outputs) folgen. In der Chemie könnte es sich um ein Gemisch handeln, das auf zugeführte Stoffe in bestimmter Weise reagiert. In der Medizin könnte es ein Patient sein, der auf Medikamente in bestimmter Weise reagiert. Allgemein nehmen wir an, dass die entsprechenden Paare von Input- und Outputdaten unabhängig durch dasselbe unbekannte Zufallsexperiment erzeugt werden. Statistisch sagt man deshalb, dass die endliche Folge von Beobachtungsdaten $(x_1, y_1), \dots, (x_n, y_n)$ mit Inputs x_i und Outputs y_i $(i = 1, \dots, n)$ durch Zufallsvariablen $(X_1, Y_1), \dots, (X_n, Y_n)$ realisiert wird, denen eine unbekannte Wahrscheinlichkeitsverteilung $P_{X,Y}$ zugrunde liegt.

Algorithmen sollen nun Eigenschaften der Wahrscheinlichkeitsverteilung $P_{X,Y}$ ableiten. Ein Beispiel wäre die Erwartungswahrscheinlichkeit, mit der für einen gegebenen Input ein entsprechender Output auftritt. Es kann sich aber auch um eine Klassifikationsaufgabe handeln: Eine Datenmenge soll auf zwei Klassen aufgeteilt werden. Mit welcher Wahrscheinlichkeit gehört ein Element der Datenmenge (Input) eher zu der einen oder anderen Klasse (Output)? Wir sprechen in diesem Fall auch von binärer Mustererkennung.

In technisch-mathematischer Sprache heißt das: Gegeben seien die Beobachtungsdaten $(x_1, y_1), \ldots, (x_n, y_n)$ mit Inputs $x_i \in \mathcal{X}$ und $y_i \in \mathcal{Y}$ $(1 \leq i \leq n)$. Die metrischen Räume $\mathcal{X}$ and $\mathcal{Y}$ sind mit einer Borel σ-Algebra versehen. Wir nehmen an, dass jedes Paar (x_i, y_i) durch dasselbe unbekannte Zufallsexperiment unabhängig erzeugt wird, d.h. es handelt sich um Realisierungen von Zufallsvariablen $(X_1, Y_1), \ldots, (X_n, Y_n)$ i.i.d. (independent and identically distributed) mit Verbundwahrscheinlichkeit $P_{X,Y}$ und Maßfunktion $X: \Omega \to \mathcal{X}$. Es geht nun darum, Eigenschaften der Verbundwahrscheinlichkeit $P_{X,Y}$ abzuleiten.

Ein Beispiel ist die Berechnung der Erwartungswahrscheinlichkeit des Outputs $f(x) = \mathbb{E}[Y|X = x]$ bei gegebenem Input x mit $\mathcal{Y} = \mathbb{R}$ (Regression). Ein weiteres Beispiel ist die binäre Klassifikation von Mustern. Beim Erkennen eines binären Musters werden die Daten einer Datenmenge $\mathcal{X}$ auf zwei mögliche Klassen verteilt, die mit $+1$ bzw. -1 bezeichnet werden. Diese Zuordnung wird durch eine Funktion $f: \mathcal{X} \to \mathcal{Y}$ mit $\mathcal{Y} = \{+1, -1\}$ beschrieben. Beim statistischen Lernen eines binären Musters geht es darum, aus einer Klasse $\mathcal{F}$ von Funktionen diejenige Zuordnung f zu ermitteln, bei der die Fehlerabweichung bzw. der erwartete Irrtum minimal ist. Wir sprechen auch von der Risikominimierung des statistischen Lernens.[10]

In technisch-mathematischer Sprache heißt das: Die binäre Musterklassifikation ordnet jedes x der Klasse mit größerer Wahrscheinlichkeit zu, also

$f(x) = \text{argmax}_{y \in \mathcal{Y}} P(Y = y|X = x)$ mit $\mathcal{Y} = \{\pm 1\}$.

Probabilistisches Lernen lässt sich als Induktionsprinzip auffassen, nach dem z.B. binäre Mustererkennung oder Klassifikation $f: \mathcal{X} \to \mathcal{Y}$ von Beobachtungen $(x_1, y_1), \ldots, (x_n, y_n)$ in die beiden Klassen $\mathcal{Y} = \{\pm 1\}$ erlernt wird. Dabei werden die Beobachtungen $(x_1, y_1), \ldots, (x_n, y_n)$ mit einer (unbekannten) Wahrscheinlichkeitsverteilung $P_{X,Y}$ (i.i.d.) erzeugt und der erwartete Irrtum (Risiko) R der Klassifikationsfunktionen f über der Funktionenklasse $\mathcal{F}$ minimiert:

$$R[f] = \int \frac{1}{2} |f(x) - y| \, dP_{X,Y}(x, y).$$

Da aber die Wahrscheinlichkeitsverteilung $P_{X,Y}$ für alle Werte unbekannt ist, kann diese Formel und damit die gesuchte Mustererkennung mit minimaler Fehlerabweichung nicht berechnet werden. Es stehen uns nur die endlich vielen empirisch beobachteten Zuordnungen $(x_1, y_1), \ldots, (x_n, y_n)$ zur Verfügung. Wir beschränken uns daher auf eine empirische Risikominimierung. Dazu ermitteln wir schrittweise für jede Zuordnungsfunktion f der Klasse $\mathcal{F}$ den empirischen Trainingsirrtum beim Lernen aus einem Sampel mit Umfang n, also

$$R_{emp}^n[f] = \frac{1}{n} \sum_{i=1}^n \frac{1}{2} |f(x_i) - y_i| \, .$$

[10] J. Peters, D. Janzing, B. Schölkopf (2017), Elements of Causal Inference. Foundations and Learning Algorithms, MIT Press: Cambridge M.A., 6f.

3.2 Anwendungsbeispiel in der Elementarteilchenphysik

Ein sensationelles Beispiel war die Entdeckung des Higgs Teilchens durch einen Machine Learning Algorithmus und Big Data. Am 28. August 2018 meldete CERN:

„Die überragende Leistung des Elementarteilchenbeschleunigers LHC (Large Hadron Collider) und moderne Machine Learning Technik ermöglichten es, die Kopplung des Higgs Bosons mit den schwersten Fermionen zu identifizieren. Damit wird erklärt, warum es Masse im Universum gibt."

Das Standardmodell der Elementarteilchenphysik erklärt, dass das Higgs Boson H in zwei Bottom Quarks b in Verbindung mit einem Z Boson zerfällt, das in Elektron e^- und ein Antielektron e^+ zerfällt. Dieses Ereignis muss unter Milliarden von Daten (Big Data) identifiziert werden, die durch Proton-Proton Kollisionen erzeugt werden. Dabei handelt es sich um ein Beispiel der Mustererkennung und Klassifikation in der Elementarteilchenphysik. Dazu müssen Signalereignisse s wie z.B. der Zerfall des Higgs-Bosons H→ $\tau^+\tau^-$ von Hintergrundereignissen h unterschieden werden.

Der Vektor $\mathbf{x} = (x_1, \ldots, x_n)$ mit n Größen eines Ereignisses (z.B. x_1 Impuls eines Leptons) wird durch eine Dichtefunktion der Verbundwahrscheinlichkeit $f(\mathbf{x}|\mathrm{s})$ für Signalereignisse und $f(\mathbf{x}|\mathrm{b})$ für Hintergrundereignisse bestimmt. Bildlich wird die Dichte der Signalereignisse mit roten Punkten und die Dichte der Hintergrundereignisse mit blauen Dreiecken dargestellt (Abb. 1). Die Selektion von Ereignismustern kann durch Schnitte (a), lineare Grenzen (b) und nichtlineare Grenzen (c) durchgeführt werden.[11]

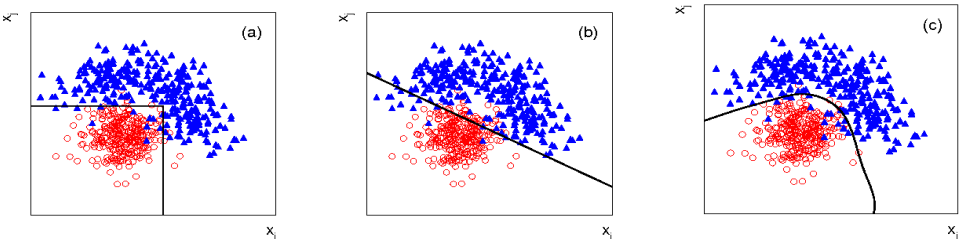


Abb. 1: Statistische Mustererkennung in der Elementarteilchenphysik (Radovic et al. 2018)

Die Konturen dieser Selektion werden durch ein konstantes Likelihood-Verhältnis der Dichtefunktionen für Signal- und Hintergrundereignisse bestimmt:

$$\lambda(\mathrm{x}) = \frac{f(\mathbf{x}|s)}{f(\mathbf{x}|b)}\,.$$

[11] A.Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, T. Wongjirad (2018), Machine learning at the energy and intensity frontiers of particle physics, in: Nature 560, 41-48.

Da die Wahrscheinlichkeitsdichten im allgemeinen unbekannt sind, ist $\lambda(\mathbf{x})$ nicht berechenbar. Allerdings könnten endliche Samples mit Trainingsdaten durch Monte Carlo Methoden berechnet werden. Ziel sind Algorithmen des Machine learning, um eine Funktion $y(\mathbf{x})$ zu finden, die am besten das Likelihood Verhältnis $\lambda(\mathbf{x})$ für die Musterselektion der Signalereignisse approximiert.

3.3 Grenzen der empirischen Induktion

Bei der Risikominimierung wird eine Folge von Funktionen der Klasse $\mathcal{F}$ mit verbessertem Trainingsirrtum erzeugt. Die zentrale Frage ist, ob durch dieses Verfahren eine Mustererkennung mit einer minimal möglichen Fehlerabweichung ermittelt werden kann. Mathematisch ist zu klären, ob die so ermittelte Funktionenfolge in der Klasse $\mathcal{F}$ gegen eine Funktion mit minimaler Fehlerabweichung konvergiert. Man spricht auch von der Konsistenz induktiver Risikominimierung, falls eine Folge von Funktionen erzeugt wird, deren Risiko gegen das kleinste mögliche Risiko konvergiert, das in der Funktionenklasse $\mathcal{F}$ mit $n \to \infty$ möglich ist.

Tatsächlich lässt sich beweisen, dass eine solche Konvergenz bzw. ein solcher Lernerfolg nur für kleine Teilklassen garantiert ist. Ein Beispiel ist die Vapnik-Chervonenkis (VC) Dimension, mit der sich die Kapazität und Größe solcher Funktionenklassen bestimmen lässt.[12] Mit großer Wahrscheinlichkeit ist dann das Risiko nicht größer als das empirische Risiko (plus einem Term, der mit der Größe der Funktionenklasse wächst.)

Bei einem Konsistenzbeweis für induktive Risikominimierung lautet die mathematische Ausgangsfrage: Konvergiert die Folge von Funktionen, die das empirische Risiko $R^n_{\mathrm{em}}[f]$ minimieren, gegen eine Funktion, die ebenfalls das Risiko $R[f]$ minimiert? Als Maß der Größe der Funktionenklasse $\mathcal{F}$ lässt sich die VC (Vapnik-Chervonenkis) Dimension heranziehen. Es lässt sich beweisen, dass für eine „kleine“ Funktionenklasse (mit kleiner VC Dimension), in der geeignete Funktionen mit kleinem empirischen Risiko $R^n_{\mathrm{em}}[f]$ vorkommen, mit hoher Wahrscheinlichkeit garantiert werden kann, dass diese Funktionen kleine erwartete Irrtümer für zukünftige Daten bei gleicher Verteilung haben. Wissenschaftstheoretisch bestimmt die VC-Dimension also die Falsifikationsmöglichkeiten einer Theorie. Damit eröffnet das statistische Lernen eine bemerkenswerte Perspektive auf Poppers Falsifikationismus, der die Dimension einer Theorie durch ihre Falsifikationsmöglichkeiten definiert hatte.[13]

Die Erfolge des Machine learning in Technik und Naturwissenschaften basieren auf großen Datensets (BIG DATA) und Rechenkapazität, mit denen diese Methoden trainiert werden können. Induktive Methoden des statistischen Lernens liefern aber nur

[12] V.N Vapnik (1998), Statistical Learning Theory, New York (N.Y.).
[13] D. Corfield, B. Schölkopf, und V. Vapnik (2009), Falsificationism and statistical learning theory: Comparing the Popper and Vapnik-Chervonenskis dimensions, in: Journal for the General Philosophy of Science 40 (1), 51-58.

Regularitäten von Wahrscheinlichkeitsverteilungen. Zugrundeliegende kausale Regularitäten erfordern zusätzliche mathematische Strukturen.

3.4 Vom statistischen zum kausalen Lernen

Datenkorrelationen können Hinweise auf kausale Zusammenhängen liefern, müssen es aber nicht. Stellen wir uns eine Testreihe vor, bei der sich eine günstige Korrelation zwischen einer verabreichten chemischen Substanz und der Bekämpfung bestimmter Krebstumore ergibt. Dann entsteht Druck des betroffenen Unternehmens, mit einem entsprechenden Medikament in die Produktion zu gehen und Gewinne abzuschöpfen. Aber auch betroffene Patienten mögen darin ihre letzte Chance sehen. Tatsächlich erhalten wir ein nachhaltiges Medikament aber nur, wenn wir den zugrunde liegen kausalen Mechanismus des Tumorwachstums, also die Gesetze der Zellbiologie und Biochemie verstanden haben.

Bereits Newton war kaum an Datenkorrelationen der fallenden Äpfel von den Apfelbäumen seines elterlichen Bauernhofs interessiert. Diese populäre Geschichte hatte Voltaire in die Welt gesetzt. Newton war vielmehr an dem zugrundeliegenden mathematischen Kausalgesetz der Gravitation interessiert, mit dem genaue Erklärungen und Prognosen sowohl der fallenden Äpfel auf der Erde als auch der Planeten am Himmel möglich wurden, letztlich auch die darauf aufbauende heutige Satelliten- und Raketentechnik. Dazu leitete er mathematische Schlüsse auf der Grundlage von Keplers Planetengesetzen und einer angenommenen Gravitationskraft ab.

Die Auseinandersetzung zwischen probabilistischem und kausalem Denken ist keineswegs neu, sondern wurde erkenntnistheoretisch bereits in der Philosophie des 18. Jahrhunderts zwischen David Hume (1711-1776) und Immanuel Kant (1724-1804) ausgefochten. Nach Hume beruht alle Erkenntnis auf sinnlichen Eindrücken (Daten), die psychologisch „assoziiert“ werden. Es gibt danach keine Kausalitätsgesetze von Ursache und Wirkung, sondern nur Assoziationen von Eindrücken (z.B. Blitz und Donner), die mit (statistischer) Häufigkeit „gewohnheitsmäßig“ korreliert werden.[14] Nach Kant sind Kausalitätsgesetze als vernunftmäßig gebildete Hypothesen möglich, die experimentell überprüft werden können. Ihre Bildung beruht also nicht auf psychologischen und statistischen Assoziationen, sondern auf der vernunftmäßigen Kategorie der Kausalität, die kausale Modelle annimmt und daraus kausale Schlüsse für Vorhersagen und Erklärungen von Beobachtungen und Wahrnehmungen ableitet.[15] Nach Kant ist dieses Verfahren seit Galileo Galilei in der Physik in Gebrauch, die so erst zur Wissenschaft wurde.

Statistisches Lernen und Schließen aus Daten reichen also nicht aus. Wir müssen vielmehr die kausalen Zusammenhänge von Ursachen und Wirkungen hinter den Messdaten

[14] D. Hume (1993), Eine Untersuchung über den menschlichen Verstand, übersetzt von R. Richter, hrsg. von J. Kulenkampff. 12. Aufl., Hamburg, 95.
[15] I. Kant (1900ff.). Hrsg.: Bd. 1-22 Preussische Akademie der Wissenschaften, Bd. 23 Deutsche Akademie der Wissenschaften zu Berlin, ab Bd. 24 Akademie der Wissenschaften zu Göttingen, Berlin, AA III, 93– KrV B 106.

erkennen. Diese kausalen Zusammenhänge hängen von den Modellen der jeweiligen Anwendungsdomäne unserer Forschungsmethoden ab, also den angenommenen Gesetzen (Gesetzeshypothesen) der Physik im Beispiel von Newton, den Gesetzeshypothesen des Standardmodells der Elementarteilchenphysik, aus denen der theoretische Physiker Higgs die Existenz des Higgs-Teilchens gefolgert hatte, den angenommenen Gesetzen der Biochemie und des Zellwachstums im Beispiel der Krebsforschung, etc. Wäre es anders, könnten wir mit den Methoden des statistischen Lernens und Schließen bereits die Probleme dieser Welt lösen. Tatsächlich scheinen das einige kurzsichtige Zeitgenossen beim derzeitigen Hype des statistischen Machine Learning zu glauben.

Statistisches Lernen und Schließen ohne kausales Domänenwissen ist aber blind – bei noch so großer Datenmenge (Big Data) und Rechenpower. Neben der Statistik der Daten bedarf es zusätzlicher Gesetzes- und Strukturannahmen der Anwendungsdomänen, die durch Experimente und Interventionen überprüft werden. Kausale Erklärungsmodelle (z.B. das Planetenmodell Newtons, das Standardmodell von Higgs oder ein Tumormodell) erfüllen die Gesetzes- und Strukturannahmen einer Theorie (z.B. die Gravitationstheorie, die Elementarteilchenphysik oder die Gesetze der Zellbiologie).

Beim kausalen Schließen werden Eigenschaften von Daten und Beobachtungen aus angenommenen Kausalmodellen, d.h. Gesetzesannahmen von Ursachen und Wirkungen, abgeleitet. Kausales Schließen ermöglicht damit, die Wirkungen von Interventionen oder Datenveränderungen (z.B. durch Experimente) zu bestimmen. Kausales Lernen versucht umgekehrt, ein Kausalmodell aus Beobachtungen, Messdaten und Interventionen (z.B. Experimente) abzuleiten, die zusätzliche Gesetzes- und Strukturannahmen voraussetzen.

Ein strukturelles Kausalmodell besteht aus einem System von strukturellen Zuordnungen von Ursachen zu Wirkungen mit eventuellen Störvariablen. Ursachen und Wirkungen werden durch Zufallsvariablen beschrieben. Ihre funktionalen Zuordnungen (unter Berücksichtigung von Störvariablen) werden durch Gleichungen definiert, also z.B. Wirkung $X_j = f(X_i, N)$ in funktioneller Abhängigkeit von Ursache X_i und Störvariable N. Anschaulich kann das Netzwerk der Ursachen und Wirkungen durch einen Graphen von Knoten und Kanten dargestellt werden. Zufallsvariablen von Ursachen und Wirkungen entsprechen Knoten. Kausale Wirkungen entsprechen gerichteten Pfeifen: $X_i \rightarrow X_j$ bedeutet, dass Ursache X_i Wirkung X_j auslöst.

Es lässt sich beweisen, dass ein Kausalmodell eine eindeutige Wahrscheinlichkeitsverteilung der Daten einschließt, aber nicht umgekehrt: Für Kausalmodelle (z.B. Planetenmodell) müssen zusätzliche Gesetze (z.B. Gravitationsgesetz) angenommen werden.[16] Um kausale Abhängigkeiten und Unabhängigkeiten von Ereignissen zu erkennen, muss die Abhängigkeit und Unabhängigkeit der sie darstellenden Zufallsvariablen ermittelt werden. Statistisch lässt sich die Unabhängigkeit der Resultate x und y zweier Zufallsvariablen (anschaulich Zufallsexperimente) X und Y dadurch ausdrücken, dass ihre Verbundwahrscheinlichkeit $p(x, y)$ faktorisierbar ist, d.h. $p(x, y) = p(x)p(y)$. Man spricht in diesem

[16] J.M. Mooij/ D. Janzing/ B. Schölkopf (2013), From ordinary differential equations to structural causal models: The deterministic case. Proceedings of the 29th Annual Conference on Uncertainty in ArtificiaL Intelligence (UAI), 440-448.

Fall auch von der Markov-Bedingung. Auf dieser Grundlage lässt sich ein Kalkül einer kausalen Unabhängigkeitsrelation ⫫ einführen.[17]

Die Kausalstruktur, die einem statistischen Modell unterliegt, kann für eine erfolgreiche Anwendung des Machine Learning folgenreich sein. So kann eine gegebene Kausalstruktur genutzt werden, um das systematische Rauschen von Messdaten bei Voraussagen von Ereignissen in der Astronomie zu reduzieren. Ein Beispiel ist das Kausalmodell, das bei der Voraussage von Exoplaneten zugrunde gelegt wird.[18] Ein Kausalmodell der klassischen Physik kann auch widerlegt werden, wie das Beispiel der quantenphysikalischen Messungen der EPR (Einstein-Podolsky-Rosen) Korrelationen zeigt. Tatsächlich bestätigt sich dabei ein neuartiges quantenphysikalisches Modell, das als „Verschränkung" (Entanglement) bezeichnet wird und von erheblicher technischer Bedeutung für Quantenkommunikation und Quantenkryptologie ist.[19]

3.5 Empirische Induktion und Bayessche Wissenschaftstheorie

Nach dem englischen Mathematiker und Theologen Thomas Bayes (1702-1762) lässt sich Lernen mit bedingten Wahrscheinlichkeiten von Ereignissen erklären.[20] Dabei wird Wahrscheinlichkeit nicht als Häufigkeit (objektive Wahrscheinlichkeit), sondern als Glaubensgrad (subjektive Wahrscheinlichkeit) von z.B. Wissenschaftler*nnen aufgefasst: Ein Ereignis A wird vor Eintreten von Ereignis B mit der a priori Wahrscheinlichkeit $P(A)$ eingeschätzt, aber nach Eintreten von B mit der a posteriori (bedingten) Wahrscheinlichkeit $P(A|B)$.

Mit Hilfe des Satzes von Bayes lassen sich bedingte Wahrscheinlichkeiten berechnen: Die bedingte Wahrscheinlichkeit $P(A|B)$ von Ereignis A nach Eintreten von Ereignis B ist durch den Quotienten der Wahrscheinlichkeit $P(A \cap B)$ (d.h. der Wahrscheinlichkeit, dass die Ereignisse A und B zusammen eintreten) und der Wahrscheinlichkeit $P(B)$ von Ereignis B definiert, d.h.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$= \frac{\frac{P(A \cap B)}{P(A)} \cdot P(A)}{P(B)} = \frac{P(B|A) \cdot P(A)}{P(B)}$$

[17] J. Pearl (2009), Causality: Models, Reasoning, and Inference, Cambridge (Mass.).
[18] J. Peters/ D. Janzing/ B. Schölkopf (2017), Elements of Causal Inference. Foundations and Learning Algorithms, Cambridge (Mass.), 158.
[19] K. Mainzer (2020), Quantencomputer. Von der Quantenwelt zur künstlichen Intelligenz, Springer: Berlin, Kap. 3, 6, 7.
[20] K. Mainzer, P. Schröder-Heister, Bayessches Theorem, in: J. Mittelstrass (Hrsg.), Enzyklopädie Philosophie und Wissenschaftstheorie Bd. 1, J.B. Metzler: Stuttgart 1995, 254-256

Daher besagt der Satz von Bayes $P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$

d.h. die Wahrscheinlichkeit von A nach Eintreten von B berechnet sich aus der bedingten Wahrscheinlichkeit von B unter Voraussetzung von A und den a priori Wahrscheinlichkeiten $P(A)$ und $P(B)$.

Wie lassen sich Bayes' Methoden auf induktive Schlüsse in den empirischen Wissenschaften anwenden? Wie lässt sich ein Modell M aus einer Datenmenge D ableiten? Aus Bayes' Theorem folgt für die Wahrscheinlichkeit eines Modells M unter Voraussetzung einer Datenmenge D:

$$P(M|D) = \frac{P(D|M)\cdot P(M)}{P(D)} = P(M)\frac{P(D|M)}{P(D)}$$

Die apriorische Wahrscheinlichkeit $P(M)$ schätzt die Wahrscheinlichkeit ein, dass Modell M korrekt ist, bevor (lateinisch: a priori) Daten vorliegen. Die aposteriorische Wahrscheinlichkeit $P(M|D)$ berücksichtigt die Einschätzung der Wahrscheinlichkeit, dass Modell M korrekt ist, nachdem (lateinisch: a posteriori) die Daten der Datenmenge D beobachtet wurden. Die Wahrscheinlichkeit $P(D|M)$, dass die Beobachtungen der Datenmenge D unter der Annahme von Modell M eintreffen, wird als Mutmaßlichkeit bzw. Plausibilität (englisch: likelihood) von Daten unter Voraussetzung eines Modells bezeichnet. Die Wahrscheinlichkeit $P(D)$ steht für die Datenevidenz der Datenmenge D. Für die Berechnung der Wahrscheinlichkeit eines Modells M unter Voraussetzung einer Datenmenge ist die Berechnung entsprechender Logarithmen häufig einfacher:

$$\log P(M|D) = \log P(D|M) + \log P(M) - \log P(D)$$

Bei der Bayesschen Einschätzung von Modellen wird gelegentlich die Voraussetzung „apriorischer" Wahrscheinlichkeiten als subjektiv oder willkürlich kritisiert. Tatsächlich nehmen aber die Effekte der apriorischen Wahrscheinlichkeiten mit wachsender Anzahl der Daten ab. Bei der logarithmischen Berechnung der Datenplausibilität (Likelihood) wächst $\log P(D|M)$ typischerweise linear mit der Anzahl der Daten aus D, während der apriorische Ausdruck $\log P(M)$ konstant bleibt. Zudem zwingt der Bayessche Ansatz, die apriorischen und aposteriorischen Annahmen klar zu unterscheiden. Die apriorische Wahrscheinlichkeit eines Modells hängt von der angenommenen Wahrscheinlichkeitsverteilung ab. Je nach Anwendungsfall eignen sich dafür z.B. Gauß- oder Dirichlet-Verteilung. Jedenfalls ist kein allgemeines objektives Prinzip bekannt, um apriorische Annahmen in allen Situationen zu bestimmen.[21]

Modelle $M(w)$ hängen von Parametern w ab. In einem physikalischen Modell von Phasenübergängen, das den Übergang von z.B. flüssigem Zustand in einen Gaszustand bei kritischen Temperaturwerten beschreibt, ist Temperatur ein Parameter. Im Unterschied zu solchen einfachen physikalischen Modellen hängen molekularbiologische Modelle von einer gewaltigen Anzahl von Parametern ab. Um einzuschätzen, wie gut ein Modell $M(w)$ für eine Datenmenge D wie z.B. eine molekularbiologische Sequenz ist, werden Fehler-

[21] J. Skilling, S. Sibisi (Eds.), Maximum Entropy and Bayesian Methods, Kluwer: Dordrecht 1996

und Irrtumsfunktionen $f(w, D) \geq 0$ angewendet, um zu berechnen, bis zu welchem Grad das Modell auf die Datenmenge in Abhängigkeit von den Modellparametern zutrifft. Es lässt sich beweisen, dass die Minimalisierung der Irrtumsfunktion äquivalent mit der maximalen Datenplausibilität (Likelihood) ist.[22]

Zwei Modelle M_1 und M_2 können verglichen werden, indem ihre Wahrheinlichkeit $P(M_1|D)$ und $P(M_2|D)$ für eine Datenmenge D (z.B. Aminosäuresequenz) verglichen wird. Ein Ziel könnte darin bestehen, das beste Modell M einer Modellklasse zu bestimmen, indem die Menge von Paramtern w mit maximaler Wahrscheinlichkeit $P(M|D)$gefunden wird. Der Bayessche Ansatz hat den wissenschaftstheoretischen Vorteil, dass er weniger komplexe Modelle bevorzugt. Das entspricht der methodischen Devise von Ockhams Rasiermesser, wonach diejenige Erklärung einer Beobachtung zu bevorzugen ist, die weniger theoretische Voraussetzungen macht.[23] Es zeigt sich nämlich, dass die Datenplausibilität (Likelihood) für gegebene Datenmengen im Schnitt kleiner wird, wenn $P(D|M)$ einen wachsenden Datenraum betrifft. Komplexe Modelle tendieren also zu einer kleineren Datenplausibilität (Likelihood) der beobachteten Daten.

Nach dem Bayesschen Theorem kommt es zur Bestimmung der Wahrscheinlichkeit $P(M|D)$ eines Modells unter Voraussetzung der Datenmenge D zunächst darauf an, die Datenplausibilität (Likelihood) $P(D|M)$ eines Modells und seine apriorische Wahrscheinlichkeit $P(M)$ einzuschätzen. Für eine Bitsequenz aus den Bits 0 und 1 könnte das simple Modell einer fairen Münze gewählt werden, deren beiden Seiten die Ziffern 0 oder 1 zeigen. Dieses Modell hat nur einen einzigen Parameter p. Die Daten bestehen aus Sequenzen über dem Alphabet $\mathcal{A} = \{0, 1\}$, die durch zufällige Münzwürfe erzeugt werden. Die Daten D einer DNA-Sequenz werden über dem Alphabet $\mathcal{A} = \{A, C, G, T\}$ gebildet. Ein einfaches Zufallsmodell wäre ein vierseitiger Würfel, dessen Seiten die Symbole A, C, G, T tragen.[24]

Die Bayessche Wissenschaftstheorie erklärt also statistisch, wie sich unsere Modelle und Hypothesen über die Welt durch neue Erfahrungen verändern und wie wir aus Erfahrung lernen können. Sie kann daher als Rahmentheorie für Lernalgorithmen verstanden werden, mit denen Lernverfahren im Machine learning automatisiert werden.

3.6 Bayessche Wissenschaftstheorie und neuronale Netze

Im Machine learning spielen neuronale Netze nach dem Vorbild des menschlichen Gehirns eine dominante Rolle. Der Durchbruch der KI-Forschung in der Praxis hängt wesentlich mit der Fähigkeit neuronaler Netze zusammen, große Datenmengen (Big Data) z.B. bei der Mustererkennung mit effektiven Lernalgorithmen anzuwenden. Praktische Anwendungen erfordern Tausende von Neuronen und Synapsen in mehrschichtigen

[22] P. Baldi, S. Brunak, Bioinformatics. The Machine Learning Approach, MIT Press: Cambridge MA 2001, 57
[23] W.H. Jeffreys, J.O. Berger, Ockham's razor and Bayesian analysis, in: Americian Science 80 1992, 64-72
[24] P. Baldi, S. Brunak, Bioinformatics. The Machine Learning Approach, MIT Press: Cambridge MA 2001, 68

neuronalen Netzen (deep learning), die statistisch mit endlich vielen Datensätzen von Inputs und Outputs trainiert werden.

Vom Bayesschen Standpunkt aus lassen sich neuronale Netze als graphische Modelle $M(w)$ mit bestimmten Parametern w auffassen. Diese Modelle erinnern vereinfacht an das menschliche Gehirn. In klassischen vorwärtslaufenden (feed-forward) neuronalen Netzen sind Neuronen als Knoten eines Graphen in Schichten angeordnet. In Abb. 2 ist jedes Neuron einer Schicht mit allen Neuronen der nachfolgenden Schicht durch gerichtete Kanten (Synapsen) verbunden, die Neuronen einer Schicht untereinander aber nicht. Ausnahmen sind die Input-Schicht, deren Neuronen keine einlaufenden Verbindungen besitzen, und die Output-Schicht, deren Neuronen keine auslaufenden Verbindungen haben. Die Schichten zwischen Input- und Output-Schicht heißen „versteckt“ (hidden). Jede Verbindung/Kante ist im graphischen Modell eines neuronalen Netzes mit einer Zahl gewichtet, die der Intensität der synaptischen Verbindung entspricht. Diese Gewichte (englisch: weights) sind die Parameter w des graphischen Modells $M(w)$ eines neuronalen Netzes. Jedes Neuron ist durch eine Aktivierungsfunktion charakterisiert, mit der die Input-Output Relation für dieses Neuron definiert wird. In Analogie zum Gehirn sagt man, dass ein Neuron „feuert“ bzw. erregt ist, wenn die Summe der gewichteten Inputs seiner Nachbarzellen einen bestimmten Schwellenwert überschreitet.

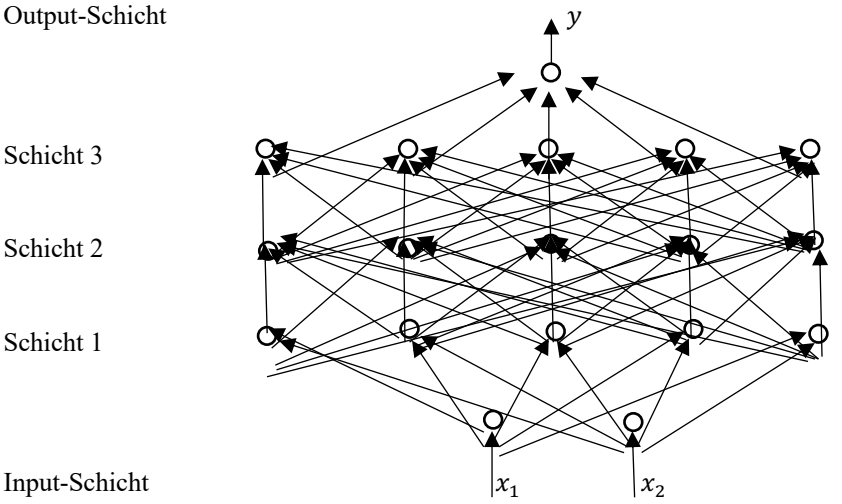


Abb.2: Vorwärtslaufendes (feedforward) Netzwerk mit 3 versteckten Schichten und einer Input- und Output-Schicht

Mathematisch können diese Netzwerke als Funktionen $v: I^n \to O^m$ definiert werden, die einen n-dimensionalen Inputbereich $I^n(n > 0)$ auf einen m-dimensionalen Outputbereich $O^m(m > 0)$ abbilden. Ein solches Netzwerk kann beispielsweise eine Approximation

einer Funktion $f: \mathbb{R}^n \to \mathbb{R}$ mit $I = O = \mathbb{R}$ berechnen. Ein Netzwerk, das z.B. 8-bit Bilder der Größe $h \times v$ (mit h horizontal und v vertikal) in zwei Klassen klassifiziert, kann durch eine Funktion $v: I^{h \cdot v} \to O$ mit Inputbereich $I = \{0, \dots, 255\}$ für die $2^8 = 256$ möglichen 8-bit Bilder und Outputbereich $O = \{0, 1\}$ für die zwei mit 0 und 1 bezeichneten Klassen definiert werden.

Die Abbildung beginnt mit einem Input aus I^n, der im Netz zunächst der Input-Schicht eingegeben und dann über die versteckten Zwischenschichten bis zur Output-Schicht weiterverarbeitet wird. Dabei werden Schicht für Schicht Linearkombinationen der Werte von Knoten (Neuronen) und Gewichten (synaptische Verbindungen) aus den vorausgehenden Schichten berechnet. Auf diese Ergebnisse werden Aktivierungsfunktionen für die nachfolgenden Neuronen angewendet (Abb.2).

Entscheidend ist, dass neuronale Netze aus Beispielen lernen können. Vom Bayesschen Standpunkt ist das nichts anderes als die Abschätzung der Tauglichkeit eines Modells $M(w)$ zur Erklärung (Fitting) einer Datenmenge D und der Abschätzung der Modellparameter w. Je nach Anwendung kann eine Datenmenge als Input-Output Sample $D = (D_1, \dots, D_K)$ aus Paaren $D_i = (d_i, z_i)$ verstanden werden, wobei d_i für die jeweiligen Daten steht und z_i für den jeweiligen Zielzustand.

Bei einer Klassifikationsaufgabe könnte es z.B. darum gehen, die Daten einer Datenmenge in bestimmte Klassen als Zielzustände einzuteilen. Wenn endlich viele Zielzustände in den Trainingsdaten vorgegeben sind, spricht man vom überwachten Lernen. Andernfalls handelt es sich um nicht-überwachtes Lernen. Durch eine Irrtumsfunktion werden die Outputdaten mit den Zielzuständen verglichen, um die Modellparameter, also die Gewichte des neuronalen Netzes zu optimieren. Vom Bayesschen Standpunkt aus handelt es sich dabei um das klassische Induktionsproblem, wie man aus Daten ein möglichst optimales Modell gewinnen kann.

Eine bemerkenswerte Eigenschaft von neuronalen Netzen ist, dass sie nahezu jede Funktion $f(x) = y$ mit beliebiger Genauigkeit approximieren können.[25] Dabei sind die Funktionsargumente x die Inputs und der Funktionswert y der Output eines passenden neuronalen Netzes. Der Algorithmus zur Approximation der Funktion f wird schrittweise über die Schichten einer geeigneten Netzwerkarchitektur realisiert. Die Bayessche Wissenschaftstheorie kann verwendet werden, um eine geeignete Zielfunktion und eine Transferfunktion für den Output auszuwählen. Dabei seien Daten wieder als unabhängige Input-Outputpaare $D_i = (d_i, z_i)$ gegeben. Diese Daten sind in dem Sinn verrauscht, dass für einen gegebenen Datenpunkt d_i verschiedene Zielzustände z_i beobachtet werden können. Die Operationen des neuronalen Netzes selbst werden als deterministisch angenommen. Für die Wahrscheinlichkeit der Datenpaare $D_i = (d_i, z_i)$ unter Voraussetzung der synaptischen Gewichte w (als Parameterwerte des neuronalen Modells $M(w)$) ergibt sich dann im Bayesschen Kalkül

[25] K. Hornik, M. Stinchcombe, H. White, Universal approximation of an unknown function and its derivatives using multilayer feedforward networks, in: Neural Networks 3 1990, 551-560

$$P\big((d_i, z_i)\big|w\big) = P(d_i|w)P(z_i|d_i, w) = P(d_i)P(z_i|d_i, w),$$

wobei für die zweite Gleichheit die Unabhängigkeit der Inputs d von den Parametern w angenommen wird. Mit der logarithmischen Formel des Bayesschen Kalküls lässt sich dann die Wahrscheinlichkeit der Modellparameter w (also der Gewichte eines neuronalen Netzes) unter Voraussetzung der Datenpaare D berechnen:

$$\begin{aligned}\log P(w|D) &= \log P(D|w) + \log P\,(w) - \log P(D)\\ &= \textstyle\sum_{i=1}^{K} \log P\big((d_i, z_i)\big|w\big) + \log P\,(w) - \log P(D)\\ &= \textstyle\sum_{i=1}^{K} (\log P(\,z_i|d_i, w) + \log P(d_i)) + \log P\,(w) - \log P(D)\\ &= \textstyle\sum_{i=1}^{K} \log P(z_i|d_i, w) + \sum_{i=1}^{K} \log P(\,d_i) + \log P\,(w) - \log P(D),\end{aligned}$$

wobei in der 3. Zeile $P\big((d_i, z_i)\big|w\big) = P(d_i)P(z_i|d_i, w)$ berücksichtigt wurde.

Bei der Berechnung kann die Datenevidenz $P(D)$ und $P(d_i)$ vernachlässigt werden, da diese Werte nicht von den Parametern w (also den synaptischen Gewichten des neuronalen Netzes) abhängen. Im Zentrum steht die Bestimmung der apriorischen Wahrscheinlichkeit $P(w)$ und der Datenplausibilität (Likelihood) $P(z_i|d_i, w)$. Bei der Datenplausibilität (Likelihood) ist die Leitidee, dass ein Netzwerk mit gegebenen Gewichten w für einen gegebenen Input d_i einen geschätzten Output $y(d_i)$ generiert. Das Modell ist vollständig definiert, wenn bestimmt werden kann, wie die beobachteten Daten $z_i = z(d_i)$ statistisch vom Output $y_i = y(d_i)$ abweichen können.

Wissenschaftshistorisch ist bemerkenswert, dass die Entwicklung neuronaler Netze mit einer scharfen Diskussion um Grenzen dieser Technik verbunden war, da sich der Lernalgorithmus des Perzeptron-Modells (1950) als linear begrenzt herausstellte. Die Lösung des Problems wurde durch die Architektur natürlicher Gehirne inspiriert. Warum sollte Informationsverarbeitung nur in einer Richtung durch die überlagerten Schichten vernetzter Neuronen laufen? D. E. Rumelhart, G. E. Hinton und R. J. Williams bewiesen 1986, dass rückgekoppelte Informationsläufe (backpropagation) zwischen Input-, Zwischen- und Outputschicht mit geeigneten Aktivierungs- und Lernalgorithmen auch nichtlineare Klassifikationen zulassen. Schließlich bewiesen K. Hornik, M. Stinchcome und H. White 1989, dass unter geeigneten Bedingungen auch Feedforward-Architekturen verwendet werden können.[26]

3.7 Anwendungsbeispiel neuronaler Netze in der Biochemie

Mit den so erweiterten neuronalen Netzen gelangen schließlich wesentliche Durchbrüche des Machine learning und der künstlichen Intelligenz. So entwickelte Google vor einigen Jahren die Software AlphaGo, dessen neuronales Netz in dem asiatischen Brettspiel Go

[26] D.E. Rummelhart, G.E. Hinton, R.J. Williams (1986), Learning representation by back-propagating errors, in: Nature 323, 533-536; K. Hornik, M. Stinchcombe, H. White (1989), Multilayer feedforward networks are universal approximators neural networks, in: Neural Networks 2, 359-366.

aus Spielerfahrung lernte und schließlich menschliche Champions schlug. Im Anschluss an diesen Erfolg wurde 2018 von derselben Firma eine Software entwickelt, deren neuronales Netz die Modellierung von Proteinen in bis dahin größter Anzahl und in kürzester Zeit gelang.[27]

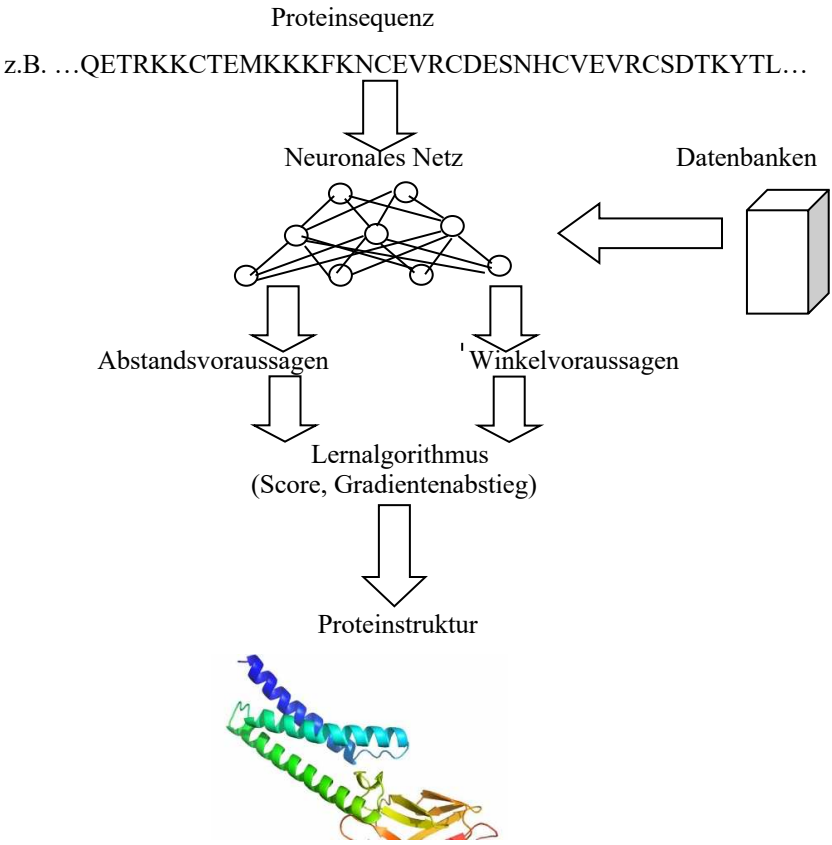


Abb. 3: Architektur von AlphaFold

Diese AlphaFold genannte Software beruht auf einem mehrschichten neuronalen Netz im Sinn von Deep Learning, das aufgrund eingegebener Sequenzen von Aminosäuren passende Formen und Faltungen von Proteinen voraussagen kann. Dazu werden Abstände und Winkel der Bindungen zwischen Aminosäuren geschätzt, deren Verteilung durch Lernalgorithmen berechnet wird. Diese Wahrscheinlichkeiten werden in einem Score zusammengefasst, mit dem abgeschätzt werden kann, wie genau eine vorgeschlagene

[27] A.W. Senior et al., Improved protein structure prediction using potentials from deep learning, in: Nature 577 2020, 706-710.

Proteinstruktur ist. Das Training des neuronalen Netzes greift auf eine umfangreiche Datenbank zurück. In Abb. 3 ist die Architektur von AlphaFold gezeigt, mit der aus einer Proteinsequenz aus den Kodes von Aminosäuren die passende Proteinstruktur erkannt wird. Der Lernalgorithmus, der aus den Wahrscheinlichkeitsverteilungen der Abstände und Winkel die passende Proteinstruktur erkennt, kann auf einem Scoring- oder einem Gradientenverfahren beruhen. Gradientenverfahren sind aus der Materialforschung bekannt, wenn durch Abkühlung (Gradientenabstieg) eines heißen Materials (z.B. Lava) bestimmte Strukturen (wie z.B. Basaltsäulen) erzeugt werden.

Bisher ist nur etwa die Hälfte aller möglichen Proteinstrukturen der menschlichen Zelle entschlüsselt. Von zentralem Interesse sind Veränderungen aufgrund von Mutationen. Missbildungen der Gestalt führen zu Fehlfunktionen als Ursachen von Krankeiten. AlphaFold und ähnliche neuronale Netze des Machine learning werden für Krankheitsbekämpfung und Gesundheitsfürsorge unverzichtbar sein, da die Diversität des Lebens auf der Komplexität der Welt der Proteine beruht. Deren Kodes lassen sich nur noch durch maschinelles Lernen mit der Rechenkapazität von Supercomputern erfassen. Grundsätzlich ist der Suchraum für mögliche Proteinformen exponentiell und nie abgeschlossen, da die Evolution nicht abgeschlossen ist.

3.8 Anwendungsbeispiel neuronaler Netze in der Evolutionstheorie

Die komplexen Strukturen und Funktionen von Genen, Proteinen und Zellen, die wir heute beobachten, sind das Ergebnis einer Evolution, die sich über viele Millionen Jahre abspielte und ständig weiterentwickelt. Darwin hatte die geniale Idee, dass sich die beobachtete Vielfalt der Arten auf gemeinsame Vorfahren zurückführen lässt, deren Entwicklungsverzweigungen sich mit den Verästlungen eines Baumes darstellen lassen. Während der Baum mit seinen Ästen und Verzweigungen ein Modell ist, entsprechen seine „Blätter“ an den Enden der Äste den beobachteten Arten. Die große Frage lautet, wie sich diese stammesgeschichtliche (phylogenetische) Entwicklung aus der molekularbiologischen Evolution ableiten lässt. In der Sprache der Bioinformatik geht es darum, die phylogenetischen Baumstrukturen aus DNA- und Proteinsequenzen abzuleiten. Dabei werden Methoden des Machine learning eingesetzt, um die großen Datenmengen mit leistungsfähigen Algorithmen und Rechnern zu bearbeiten.

Mathematisch wird unter einem Baum T ein verbundener azyklischer Graph verstanden (Abb. 4). In einem Baum sind zwei Knoten immer mit genau einer Kante verbunden. Die Anzahl der Knoten ist immer um genau 1 größer als die Anzahl der Kanten. Ein Baum heißt binär, falls jeder Knoten entweder einen oder drei Nachbarknoten hat. Mit r (englisch: root) wird der ausgezeichnete Wurzelknoten eines Baums bezeichnet. Man unterscheidet verwurzelte und nicht verwurzelte Bäume ohne ausgezeichneten Wurzelknoten. In phylogentischen Bäumen repräsentiert der Wurzelknoten die Voräufersequenz, von der alle anderen Sequenzen des Baums abstammen bzw. abgeleitet werden können. Charakteristisch für Bäume sind ihre Topologie und die Länge ihrer Äste. Die Topologie zeigt das Verzweigungsmuster und Divergenz der Evolution. Die Länge der Äste gibt den

zeitlichen Abstand zwischen den Ereignissen an, die durch die jeweiligen Sequenzen repräsentiert werden.

Als ein vereinfachtes probabilistisches Modell der Evolution kann wieder ein Würfelmodell angenommen werden. Danach startet die Evolution mit der Sequenz eines Vorfahrens und entwickelt sich schrittweise, indem Buchstaben zufällig und unabhängig von der Position ersetzt werden. In Abb. 4 ist r die Baumwurzel. Der zeitliche Abstand zwischen den Baumknoten i und j wird mit d_{ij} bezeichnet. X_i ist der Buchstabe an dem verborgenen Baumknoten i .Die beobachteten Buchstaben befinden sich an den unteren Knoten, die Blätter genannt werden. Die Wahrscheinlichkeit für die Ersetzung von Buchstaben X_i durch Buchstaben X_j beim Übergang von Knoten i nach j wird mit $p_{X_jX_i}(d_{ji})$ bezeichnet.[28]

Ein einfaches probabilistisches Modell für die Evolution geht von folgenden Annahmen aus:

- An jeder Stelle gibt es nur Ersetzungen von Buchstaben (also keine Einsetzung und Löschung). Alle beobachteten Sequenzen haben die gleiche Länge.
- Ersetzungen an jeder Stelle sind unabhängig voneinander.
- Ersetzungswahrscheinlichkeiten hängen nur vom laufenden (gegenwärtigen) Zustand ab und nicht von der vergangenen Entwicklungsgeschichte (Markov-Bedingung)
- Der Markov-Prozess ist für alle Positionen gleich.

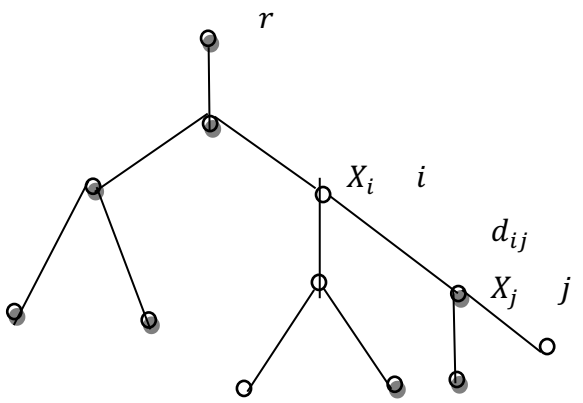


Abb. 4: Graphisches Modell eines binären phylogenetischen Baums

[28] P. Baldi, S. Brunak (2001), Bioinformatics. The Machine Learning Approach, MIT Press: Cambridge MA, 267.

Der Nachteil dieser vereinfachten Annahmen ist, dass in biologischen Evolutionen keine einzige Annahme von der realen DNA erfüllt wird. Die Länge von DNA-Sequenzen kann sich aufgrund von Einsetzungen und Löschungen verändern. Ferner ist die Evolution von verschiedenen Positionen nicht unabhängig. Die Veränderungsraten während der Evolution sind zeitlich und mit Bezug auf die Positionen nicht uniform. Zudem kann die DNA rekombiniert werden. Entsprechend vereinfachte probabilistische Modelle können also nur als Approximationen dienen.

Wenn eine Menge von Sequenzen und ein probabilistisches Evolutionsmodell gegeben sind, kann versucht werden, die wahrscheinlichste Baumtopologie und die wahrscheinlichste Länge der Äste zu bestimmen.[29] Über einem Alphabet $\mathcal{A}$ seien K Sequenzen $O_1, \dots, O_K$ derselben Länge N gegeben. Im entsprechenden Baummodell T wird mit r die gemeinsame Wurzel und mit d_{ij} der Abstand zwischen benachbarten Knoten i und j bezeichnet. Ziel der Bayesschen Wissenschaftstheorie ist es, die Wahrscheinlichkeit $P(T|O_1, \dots, O_K)$ des Modells T bei Kenntnis der Sequenzen $O_1, \dots, O_K$ zu bestimmen. Nach dem Bayesschen Theorem ist ein wesentlicher Schritt dazu die Berechnung der Plausibilität (Liklihood) $P(O_1, \dots, O_K|T)$ dieser Sequenzen für das probabilistische Evolutionmodell T. Wegen der Unabhängigkeitsannahme des Evolutionsmodells lässt sich diese Wahrscheinlichkeit für die einzelnen Buchstaben der Sequenzen faktorisieren:

$$P(O_1, \dots, O_K|T) = \prod_{k=1}^{N} P(O_1^k, \dots, O_K^k|T) \,,$$

wobei O_j^k den k-ten Buchstaben bezeichnet, der in der j-ten Sequenz beobachtet wird. Es genügt also, die Wahrscheinlichkeiten $P(O_1^k, \dots, O_K^k|T)$ zu untersuchen, die sich auf die k-ten Buchstaben O_j^k in den Sequenzen an den K Blättern des Baums beziehen. An jedem Baumknoten i lässt sich eine versteckte Zufallsvariable χ_i annehmen, die den Buchstaben angibt, der mit Knoten i verbunden ist. Daher lässt sich ein phylogenetischer Baum auch als Kausalmodell mit Baumstruktur verstehen, in dem die bedingte Wahrscheinlichkeit von Knoten j bei gegebenem Elternknoten i von der zeitlichen Distanz d_{ji} abhängt:

$$P\big(\chi_j = \mathrm{Y}\big|\chi_i = \mathrm{X}\big) = p_{\mathrm{YX}}(d_{ji})$$

Alle bekannten Algorithmen zur Berechnung von Kausalmodellen lassen sich daher auf die phylogenetischen Baummodelle anwenden.

Die Suche nach optimalen Baumstrukturen ist eine rechenintensive Herausforderung. Der Suchraum aller möglichen Bäume ist exponentiell groß, so dass eine erschöpfende Suche ausgeschlossen ist. Es gibt also nur heuristische Verfahren. Dabei wird häufig in jedem Entwicklungsschritt eine neue Spezies unter Berücksichtigung aller ihrer möglichen Positionen im aktuellen Baum selektiert. Probabilistische Evolutionsmodelle dürfen allerdings nicht zu einfach sein. Markov-Modelle erweisen sich für langfristige Evolutionsabläufe als unrealistisch, da sie in diesem Zeitraum mathematisch zu einer Gleichgewichtsverteilung konvergieren. Ein Gleichgewichtszustand der Natur widerspricht aber allen bisherigen Erfahrungen. Gleichgewichte treten nur vorübergehend und lokal in Teilsystemen auf.

[29] J. Felsenstein, Evolutionary trees from DNA sequences: A maximum likelihood approach, in: Journal of Molecular Evolution 19 1981, 368-376

Daher liegt es nahe, verschiedene lokale Evolutionsmodelle zu kombinieren. Dabei ergeben sich hochgradig nichtlineare dynamische Systeme, die zwar der biologischen Komplexität besser entsprechen mögen, aber an Grenzen der Berechenbarkeit stoßen.

Im Zeitalter der Pandemien ist die virale Evolution von zentraler Bedeutung. Im Machine learning konzentriert sich die Forschung auf eine Modellierung der Struktur und Funktionen eines Virus. Eine Funktion besteht darin, dass Teile eines Virus durch Antikörper angegriffen werden können, um das virale Eindringen in eine Zelle und die Ausbreitung des Virus im Organismus zu unterbinden. Eine weitere Funktion betrifft Proteinfragmente eines Virus, die sich auf der Oberfläche einer menschlichen Zelle abbilden und damit die Zelle als infiziert markieren, um so von Antikörpern erkannt und ausgeschaltet werden zu können. Modelle des Machine learning wurden trainiert, um Voraussagen über den Intensitätsgrad dieser Eigenschaften für jedes virale Fragment abzuleiten. Damit lässt sich besser und schneller abschätzen, welche Teile eines Virus einen höheren Grad an Immunogenität aufweisen, also die Fähigkeit, eine Immunantwort zu entwickeln. Diese Teile können dann in einen Impfstoff eingebaut werden.

Die große Stärke des Machine learning besteht darin, Muster und Zusammenhänge in großen Datenmassen zu erkennen. Mit dieser Fähigkeit ist Machine learning menschlichen Fähigkeiten weit überlegen. In der Immunologie geht es um nahezu eine Million von Proteinfragmenten, die auf einer Zelloberfläche präsentiert und für T-Zellen sichtbar sind. Kein menschlicher Forscher wäre in der Lage, diese Aufgabe systematisch in vernünftiger Zeit für ein bestimmtes Fragment des Corona-Virus zu erledigen. Für ein geeignetes Modell könnte demgegenüber mit Algorithmen des Machine learning die Datenplausibilität von Proteinfragmenten berechnet werden, um so den besten Kandidaten vorauszusagen. Mit dieser Methode wurden Regionen in der Proteinhülle von SARS-CoV-2 gefunden, die starke Antikörperziele bilden.

Auf dieser Grundlage können im Computer Impfstoffe „designed“ werden. Dabei wird das als günstig erkannte virale Proteinfragment mit anderen virusähnlichen Partikeln zusammengelagert, damit der Impfstoff für das Immunsystem wie ein echter Virus erkannt wird, um daraufhin Antikörper zu entwickeln. Was so einfach klingt, ist in der Biologie äußerst komplex. Proteine bestehen aus zehntausenden von Molekülen. Die möglichen Faltungen von Proteinen, mit denen ihre Funktionen festgelegt werden, sind exponentiell vielfältig und rechnerisch nicht vollständig zu erfassen. Es kann zwar im Computermodell von vornherein eine große Anzahl von Möglichkeiten ausgeschlossen werden, die deshalb nicht mehr im Labor getestet werden müssen. Aber als günstig erkannte Kandidaten müssen nach wie vor im Labor getestet und verworfen werden, um dann im Modell andere Kandidaten auszuwählen, die wieder im Labor zu testen sind.

In einer Art von methodischer Spirale arbeiten also Machine learning und Labortest zusammen und approximieren ein günstiges Ergebnis. Der Wissenschaftsphilosoph Karl R. Popper hat diese Forschungsspirale so beschrieben: „Wir irren uns empor!“[30] Dieses Wort beschreibt anschaulich die hier verfolgte Forschungsstrategie. Ihre Grenze ist aber

[30] Dieses Zitat fasst Poppers Position in seinem Buch „Objektive Erkenntnis. Ein evolutionärer Entwurf“ (1973) zusammen.

eindeutig: Am Ende gibt es keinen mathematischen Beweis, dass die Forschungsspirale in jedem Fall zu einem optimalen Ergebnis oder gar zur Wahrheit konvergiert.

Die dramatischen Einbrüche, die virale Pandemien in Wirtschaft und Gesellschaft weltweit auslösen, führen zu der Frage, ob die Menschheit sich langfristig auf neuen Pandemiewellen vorbereiten kann. Die virale Evolution geht nämlich mit größerer Geschwindigkeit weiter als die Evolution der Pflanzen und Tiere. Dieser Prozess wird im Zeitalter der Globalisierung erheblich beschleunigt, da die epidemologische Ansteckungsgefahr bei großen Menschenmassen weltweit wächst und damit die Ausgangslage der viralen Evolution mit hoch vernetzten Wirtsorganismen immer günstiger wird. Kurz gesagt: Die Gefahr der Pandemien steigt und wegen der Steigerung der viralen Fitness auch ihre Gefährlichkeit. Schließlich könnte dieser Beschleunigung nur noch mit Machine Learning, großen Datenbanken und Superrechnern begegnet werden.

Nun lassen sich die exponentiell wachsenden Möglichkeiten viraler Evolutionsbäume nicht vollständig mit noch so großen Supercomputern abbilden. Denkbar wäre aber, dass für Klassen solcher viralen Evolutionsbäume unter bestimmten Einschränkungen (constraints) genetisch erzeugte Gegenmaßnahmen prognostizierbar sind. Auf diesem Weg könnte eine Art Toolbaukasten von Lernalgorithmen angelegt werden, um im Fall der Fälle schnell reagieren zu können. Das ganze Unternehmen wäre selbst ein lernendes System, das durch neue Erfahrungen schrittweise ausgebaut wird. Wir begegnen also der viralen Evolution mit einer Evolution der Künstlichen Intelligenz.

4. Künstliche Intelligenz als Modell der mathematischen Physik

Neuronale Netze sind der Schlüssel zum Erfolg des Machine Learning in Naturwissenschaften und technischer Anwendung. Allerdings kommen dort mittlerweile Netze mit Hunderten und Tausenden von Schichten und Millionen von Neuronen zum Einsatz. Mit Tiefe (depth) wird die Anzahl der neuronalen Schichten bezeichnet. Man spricht mittlerweile nicht mehr nur von Deep Learning, sondern bei dieser enormen Anzahl von Schichten auch von „Deeper Learning“. Training und Abläufe dieser Netze sind nur noch statistisch beschreibbar und erscheinen zufällig. Verifikationen ihres Verhaltens stützen sich in der Regel nur noch auf punktuelle Tests mit Trial-and-Error Verfahren. Damit verschärft sich das Black Box Problem dieser Netze mit ihrer wachsenden Tiefe und Breite in der Praxis.

Gibt es eine mathematische Theorie, um die Vorgänge in diesen großen neuronalen Netzen zu verstehen und zu erklären? Tatsächlich lassen sich neuronale Netze in der mathematischen Theorie komplexer dynamischer Systeme erklären, die in meiner Arbeit seit vielen Jahren die Leitlinie bildet. Historisches Vorbild war die statistische Mechanik, in der z.B. eine Flüssigkeit als komplexes System aus vielen Flüssigkeitsmolekülen als Elementen verstanden wird. Makroskopisches Verhalten wie Strömungsmuster werden aus den nichtlinearen Wechselwirkungen der mikroskopischen Elemente abgeleitet. Der dabei verwendete mathematische Formalismus ist aber unabhängig von der Anwendung und kann daher in Natur-, Wirtschafts- und Sozialwissenschaften zur Modellbildung

angewendet werden.[31] Wissenschaftstheoretisch handelt sich also nicht um eine „naturalistische“, sondern mathematische Methodologie, die unabhängig von Anwendungsmodellen ist.

Als mikroskopische Elemente tiefer neuronaler Netze werden künstliche Neuronen verwendet. Ihre Wechselwirkungen werden durch synaptische Gewichte repräsentiert. Makroskopische Verhaltensmuster werden durch statistische Verteilungsfunktionen beschrieben. Das Entstehen („Emergenz“) der makroskopischen Muster und Strukturen findet in Phasenübergängen statt, die von kritischen Kontrollparametern abhängen. Emergenz ist hier also keine philosophische Metapher, sondern beschreibt die mathematisch beschreibbare Entstehung makroskopischer Strukturen in Phasenübergängen komplexer dynamischer Systeme.

Abb. 2 zeigt ein vielschichtiges neuronales Netz mit Input x, der in einer Folge von Zwischensignalen $s^{(1)}$, $s^{(2)}$ und $s^{(3)}$ der Zwischenschichten in den Output $y = f(x;\theta)$ transformiert wird. Dabei repräsentieren die Parameter θ die Zahlenwerte, mit denen die Signale in den Synapsen zwischen den Neuronen in Abhängigkeit von ihren Schwellenwerten des „Feuerns“ gewichtet werden. In praktischen und kommerziellen Anwendungen kann es sich um Milliarden von Parametern handeln. Die Parameter θ sind die Kontrollparameter des komplexen Systems.

Sieht man von den Schichten ab, stellt die parametrisierte Funktion $f(x;\theta)$ das neuronale Netz als komplexes System mit Input x und Kontrollparameter θ dar. Wegen der gewaltigen Zahl von Parametern wird das Netzwerk in der Praxis mit einer Zufallsauswahl einer Wahrscheinlichkeitsverteilung $p(\theta)$ initialisiert. In einem nächsten Schritt $\theta \to \theta^*$ werden die Parameter so justiert, dass die resultierende Netzwerkfunktion $f(x;\theta^*)$ eine gewünschte Zielfunktion $f(x)$ so nahe wie möglich approximiert: $f(x;\theta^*) \approx f(x)$.

Um die Werte θ^* zu finden, wird die Netzwerfunktion $f(x;\theta)$ mit Daten $(x, f(x))$ trainiert, die bei der Zielfunktion beobachtet wurden. Es handelt sich also um den induktiven Lernprozess, der grundlegend für Machine Learning ist und in den Abschnitten 3.1 und 3.6 bereits erklärt wurde. Die Justierung der Parameter wird „Training“ genannt und das Verfahren der Justierung „Trainingsalgorithmus“.

Im Sinn komplexer dynamischer Systeme muss das makroskopische Verhalten der trainierten Netzwerkfunktion $f(x;\theta^*)$ aus den mikroskopischen Beschreibungen des Netzwerks mit den trainierten Parameterwerten θ^* erklärt werden. Um die Zielapproximation $f(x;\theta^*) \approx f(x)$ zu verstehen, muss erklärt werden, wie das trainierte Netzwerk $f(x;\theta^*)$ die Trainingsdaten $(x, f(x))$ bei der Approximation von $f(x)$ nutzt. Die praktische Herausforderung besteht in der Beherrschung der riesigen Parameterzahl.

Mathematisch wird das Netzwerk durch eine hoch nichtlineare Funktion $f(x;\theta^*)$ repräsentiert. Um sie mathematisch beherrschbar zu machen, wird ihre Taylor-Entwicklung in der Umgebung des initialisierten Werts von θ untersucht. Eine Taylorreihe wird in der

[31] K. Mainzer K (2007), Thinking in Complexity. The Computational Dynamics of Matter, Mind, and Mankind. Springer. Berlin 5. Aufl.

Analysis verwendet, um eine (glatte) Funktion in der Umgebung einer Stelle durch eine Potenzreihe darzustellen, die der Grenzwert der Taylor-Polynome ist. Polynome sind mathematisch wesentlich einfacher zu berechnen als andere Funktionen:[32]

$$f(x;\theta^*) = f(x;\theta) + (\theta^* - \theta)\frac{df}{d\theta} + \frac{1}{2}(\theta^* - \theta)^2\frac{d^2f}{d\theta^2} + \cdots$$

mit $f(x;\theta)$ und ihren Ableitungen für die initialisierten Werten der Parameter θ.

Allerdings zeigen sich auch für die Taylor-Entwicklung der Netzwerkfunktion $f(x;\theta^*)$ erheblich und grundsätzliche Probleme der Berechnung:

(1) Es muss eine unendliche Anzahl von Termen $f, \frac{df}{d\theta}, \frac{d^2f}{d\theta^2}, \ldots$ berechnet werden. Da die Differenzen $(\theta^* - \theta)$ zwischen trainierten und initialisierten Parametern groß werden, wächst die Anzahl der benötigten Terme für eine geeignete Approximation der trainierten Netzwerkfunktion $f(x;\theta^*)$.

(2) Da die Parameter θ ein zufälliges Sample der initialisierten Verteilung $p(\theta)$ sind, erhält man mit jeder Initialisierung des Netzwerks eine verschiedene Funktion $f(x;\theta)$. Jeder Term $f, \frac{df}{d\theta}, \frac{d^2f}{d\theta^2}, \ldots$ wird damit zu einer Zufallsfunktion des Inputs x. Damit führt jede Initialisierung zu einer Verteilung über der Netzwerkfunktion und ihren Ableitungen, die in einer Abbildung

(3) $p(\theta) \rightarrow p(f, \frac{df}{d\theta}, \frac{d^2f}{d\theta^2}, \ldots)$

(4) der Wahrscheinlichkeitsverteilung der initialen Parameter θ auf die Verbund-Wahrscheinlichkeitsverteilung der Netzwerkfunktion $f(x;\theta)$, ihrem Gradienten $\frac{df}{d\theta}$, ihrer Hesse-Matrix $\frac{d^2f}{d\theta^2}$ etc. dargestellt wird. Diese Verbund-Wahrscheinlichkeitsverteilung fasst also eine unendliche Anzahl von Zufallsfunktionen zusammen, die analytisch nicht beherrschbar ist.

(5) Der gelernte Wert der Parameter θ^* ist das Ergebnis eines komplizierten Trainingsprozesses. Der Lernalgorithmus erfolgt iterativ in Lernschritten, die in komplizierter Weise nichtlinear voneinander abhängen und praktisch kaum handhabbar sind.

Könnten die drei Probleme (1)-(3) gelöst werden, wäre die Taylor-Entwicklung eine geeignete Erklärung für die trainierte Netzwerfunktion $f(x;\theta^*)$. Man müsste also eine Wahrscheinlichkeitsverteilung

$$p(f^*) \equiv p(f(x;\theta^*)|\text{Lernalgorithmus}, \text{Traingsdaten})$$

[32] D.A. Roberts, S. Yaida, B. Hanin (2022), The Principles of Deep Learning Theory. An Effective Theory Approach to Understanding Neural Networks. Cambridge University Press: Cambridge., S. 5 ff.

über den trainierten Netzwerkfunktionen $f(x;\theta^*)$ in Abhängigkeit vom Lernalgorithmus und den Trainingsdaten finden.

Tiefes und tiefstes Lernen hängt von der Größe der Netzwerke ab, die durch die Tiefe des Netzes, d.h. die Anzahl L der Schichten, und seine Breite, d.h. die Anzahl der Neuronen pro Schicht bestimmt ist. Um die Verteilung $p(f^*)$ über trainierten Netzwerken zu vereinfachen, ist es naheliegend, das Verhalten von extrem großen Netzwerken in theoretischen Limits zu untersuchen. Dazu werden idealisierte Netzwerke unendlicher Breite mit

$$\lim_{n\to\infty} p(f^*)$$

angenommen. In diesem Fall ergeben sich für die Berechnung extreme Vereinfachungen:

(1) Alle höheren Ableitungsterme $\frac{d^k f}{d\theta^k}$ mit $k \geq 2$ verschwinden und es bleiben nur die linearen Terme $f, \frac{df}{d\theta}$.
(2) Die Verteilungen der Zufallsfunktionen sind unabhängig und führen zu einem sehr einfachen Ausdruck
(3) $\lim_{n\to\infty} p\left(f, \frac{df}{d\theta}, \frac{d^2 f}{d\theta^2}, \ldots\right) = p(f)p(\frac{df}{d\theta})$.
(4) Der Trainingsprozess ist linear und vollständig unabhängig von allen Details des Lernalgorithmus. Daher lässt sich eine analytische Lösung für θ^* finden. Die trainierte Verteilung $\lim_{n\to\infty} p(f^*)$ ist in diesem Fall eine einfache Gaußverteilung (mit einem von Null verschiedenen Mittel).

Leider erlaubt eine solche formale Grenzwertbetrachtung mit dem Limit unendlicher Breite der Netzwerke nur sehr simple Modelle von neuronalen Netzen. Alle detaillierten Korrelationen und genauen Details von Neuronen werden eliminiert, die in realen sehr großen, aber endlichen Netzwerken eine Rolle spielen.

Daher wird im Folgenden vorgeschlagen, den Grenzwert für Netzwerke mit unendlicher Breite so zu korrigieren, dass die Korrekturen kleiner werden, wenn die Breite größer wird. Dazu werden Ideen der Störungstheorie aufgegriffen, die in der Physik bei wechselwirkenden Systemen angewendet wird. Der Prozess des Deep Learning wird nun durch eine $1/n$-Expansion mit dem Kehrwert $\epsilon = 1/n$ der Schichtenbreite n und einem kleinen Expansionsparameter $\epsilon \ll 1$ beschrieben. Mathematisch wird dazu der Grenzwert mit unendlicher Breite durch eine Taylor-Expansion mit folgenden Korrekturtermen ergänzt:[33]

$$p(f^*) \equiv p^{(0)}(f^*) + \frac{p^{(1)}(f^*)}{n} + \frac{p^{(2)}(f^*)}{n^2} + \cdots$$

[33] D.A.Roberts, S. Yaida, B. Hanin (2022), The Principles of Deep Learning Theory. An Effective Theory Approach to Understanding Neural Networks. Cambridge University Press: Cambridge., S. 9 ff., 391-399.

mit $p^{(0)}(f^*) \equiv \lim_{n\to\infty} p(f^*)$ und Korrekturen $p^{(k)}(f^*)$ mit $k \geq 1$. Es lässt sich zeigen, dass bereits eine Berechnung des ersten Korrekturgliedes die Probleme (1)-(3) einer Taylor-Expansion beherrschbar machen. In diesem Fall wird die Taylor-Expansion bis auf die ersten beiden Terme abgeschnitten:

$$p(f^*) \equiv p^{(0)}(f^*) + \frac{p^{(1)}(f^*)}{n} + O(\frac{1}{n^2}) .$$

(1) Höhere Ableitungsterme $\frac{d^k f}{d\theta^k}$ mit $k \geq 4$ tragen Werte in der Größenordnung $1/n^2$ und kleiner bei. Um die wesentlichen Beiträge der Größenordnung $1/n$ zu berücksichtigen, genügen die vier Terme $f, \frac{df}{d\theta}, \frac{d^2 f}{d\theta^2}, \frac{d^3 f}{d\theta^3}$.

(2) Die Verteilung der Zufallsfunktionen bei der Initialisierung $p\left(f, \frac{df}{d\theta}, \frac{d^2 f}{d\theta^2}, \frac{d^3 f}{d\theta^3}\right)$ ist nahezu einfach in der Größenordnung und kann in allen Details in der Störungstheorie behandelt werden.

(3) Mit der Störungstheorie lässt sich in diesem Fall die nichtlineare Trainingsdynamik beherrschen und eine analytische Lösung für θ^* finden.

Die trainierte Verteilung der Größenordnung $1/n$ entspricht einer fast-Gauß-Verteilung, die analytisch erfasst werden kann. Die Beschränkung auf die Größenordnung $1/n$ reicht aus, um die Verteilung $p(f^*)$ über den trainierten Netzwerkfunktionen zu berechnen und zu verstehen. Details der Wechselwirkungen zwischen den Neuronen sind nun berücksichtigt. Damit entspricht dieser Ansatz deutlich mehr realen technischen neuronalen Netzen als der Ansatz mit Netzen unendlicher Breite.

Das Verhältnis von Breite n und Tiefe L eines neuronalen Netzes erlaubt eine Klassifizierung der Wirkungsweise von geschichteten Netzen. Für die Proportion $r \equiv L/n$ gilt:

- Im Grenzwert $r \to 0$ werden die Wechselwirkungen zwischen Neuronen ausgeschaltet. Netzwerke mit Grenze $\lim_{n\to\infty} p(f^*)$ unendlicher Breite sind tatsächlich nicht wirklich tief, da die relative Tiefe mit $\frac{L}{n} = 0$ Null beträgt.
- Im Intervall $0 < r \ll 1$ gibt es signifikante Wechselwirkungen zwischen Neuronen. Die Beschränkung auf Netzwerke endlicher Breite der Größenordnung $1/n$ erlaubt genaue Berechnungen des Outputs des trainierten Netzwerks. In dem Sinn sind solche Netzwerke effektiv tief.
- Im Intervall $r \gg 1$ werden die Neuronen stark verbunden. Dann verhalten sie sich chaotisch. Wegen der starken Fluktuationen ist keine effektive Berechnung mehr möglich. In dem Sinn sind diese Netzwerke „zu tief".

Für praktische und kommerzielle Anwendungen folgt daraus, dass große Netze von endlicher Breite den größten Erfolg effektiver Berechenbarkeit und Beherrschbarkeit versprechen. Es ist bemerkenswert, dass diese zentrale Einsicht für praktische Anwendungen

neuronaler Netze aus einer mathematischen Theorie der Physik folgt, in der die Analyse des Unendlichen eine zentrale Rolle spielt. Es bestätigt sich erneut: Um ein Problem vertieft zu verstehen und erklären zu können, ist eine gute mathematische Theorie unerlässlich.

References

K. Mainzer/R. Kahle, Grenzen der Künstlichen Intelligenz – theoretisch, praktisch, ethisch, Springer: Berlin 2022.

K. Mainzer, Künstliche Intelligenz. Wann übernehmen die Maschinen? Springer: Berlin 2. Auflage 2019 (englische und chinesische Übersetzung 2022).

K. Mainzer (Hrsg.), Philosophisches Handbuch der Künstlichen Intelligenz, Springer: Berlin (online: https://link.springer.com/referenceworkentry/10.1007%2F978-3-658-23715-8_51-1)

Workshop on Machine Learning for Astroparticle Physics and Astronomy (ml.astro)

Unification of Algorithms for Quantification and Unfolding

Mirko Bunse[1]

Abstract: Quantification is the supervised learning task of predicting the prevalences of classes in a data sample. Physics literature knows the same task under a different name: unfolding. However, the literature on quantification and the literature on unfolding are largely disconnected from each other. We bridge this interdisciplinary gap by proposing a common framework that integrates algorithms from both fields in a unified form. Instantiations of our framework differ from each other in terms of the loss functions, the regularizers, and the feature transformations they employ.

Keywords: Quantification; Unfolding; Classification; Experimental physics; Machine learning

1 Introduction

Many applications of supervised learning require a prediction of the *distribution* of the target quantity, as exhibited by some data sample. In these applications, predictions for individual data instances are only secondary; they are issued as a means from which the distribution can be reconstructed. Examples of such applications are the reconstruction of energy spectra in astroparticle physics [Bu18], text sentiment analyses [GS16], technical support log analyses [Fo08], social sciences [HK10], and several other areas.

Supervised learning for the prediction of target distributions is known as *quantification learning* [Fo08; Go17]. Within experimental physics, however, the same problem is called *unfolding* [Bl85; DA95; HK96] or *deconvolution* [Bu18]. As of today, the literature from quantification research and the literature from unfolding research are largely disconnected from each other, despite their substantial similarities in terms of their problem statements and their solutions.

Contributions We propose a common framework for algorithms that stem from quantification literature and from unfolding literature. This framework reveals several similarities between existing methods from the two research fields. Moreover, it paves the way for strengthening the interdisciplinary efforts on the subject. Our presentation completes a similar unification attempt by Firat [Fi16] in terms of i) taking unfolding algorithms into consideration and ii) giving formal proofs about the correctness of our framework.

Sect. 2 details unfolding algorithms within our unifying framework, in which we integrate algorithms from quantification literature in Sect. 3. We summarize our findings in Tab. 1 before Sect. 4 concludes.

[1] TU Dortmund University, Artificial Intelligence Group, D-44221 Dortmund, Germany
mirko.bunse@cs.tu-dortmund.de

doi:10.18420/inf2022_37

2 Unfolding

A frequent objective in experimental physics is to estimate the distribution of a physical quantity that cannot be measured directly. In this case, the distribution needs to be reconstructed from correlated quantities which are measured instead.

In particular, we intend to estimate the distribution $p(y) = \mathbb{P}(Y_c = y)$ of the relevant quantity Y_c, assuming that we can measure the distribution $q(\vec{x}) = \mathbb{P}(X = \vec{x})$ of some other quantity X within a sample. Moreover, let the measurement process be characterized through the conditional probabilities $M(\vec{x} \mid y_c) = \mathbb{P}(X = \vec{x} \mid Y_c = y)$ of measuring some $\vec{x} \in \mathcal{X}$ when the relevant quantity has the (possibly continuous) value $y \in \mathcal{Y}_c$. The objective of any unfolding algorithm is then to reconstruct the relevant distribution p from the distributions q and M, as according to the integral

$$q(\vec{x}) \;=\; \int_{\mathcal{Y}_c} M(\vec{x} \mid y) \cdot p(y) \,\mathrm{d}y. \tag{1}$$

The estimation of p from data is enabled through the discretization of Eq. (1). In case of a continuous target interval $\mathcal{Y}_c = [a, b)$, we first need to map each continuous label to a discrete class index $\mathcal{Y} = \{1, \ldots, C\}$. For instance, the estimation of an energy spectrum requires a binning of the interval $\mathcal{Y}_c$ into C bins [Bl02; DA95]. We proceed similarly with the feature space $\mathcal{X} \subseteq \mathbb{R}^d$, in mapping it to a discrete feature representation $f(\vec{x}) \in \{1, \ldots, F\}$, which is still to be defined for each unfolding algorithm in particular.

The discretization of y and $\vec{x}$ gives rise to a straightforward representation of distributions in terms of histograms. Consider a data sample $D = \{(\vec{x}_i, y_i) \in \mathcal{X} \times \mathcal{Y} : 1 \leq i \leq N\}$ in which the classes y_i are not observed. Estimating the quantities from Eq. (1) in terms of histograms

$$\vec{p} \;=\; \frac{1}{N} \sum_{i=1}^{N} \delta_{y_i}, \qquad \vec{q} \;=\; \frac{1}{N} \sum_{i=1}^{N} \delta_{f(\vec{x}_i)}, \qquad [\delta_j]_k \;=\; \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

leads to the system of linear equations

$$\vec{q} \;=\; M \cdot \vec{p}, \tag{3}$$

where the transfer matrix $M \in \mathbb{R}^{C \times F}$ is estimated by counting and normalizing the co-occurrences of labels y and transformed features $f(\vec{x})$ in a training set. Advanced algorithms are required to estimate $\vec{p}$ because a direct solution $M^{-1}\vec{q}$ typically leads to estimates that are not physically plausible.

A Common Framework for Unfolding and Quantification Unfolding algorithms solve Eq. (3) for $\vec{p}$, a histogram estimate of the (continuous) distribution $p(y)$ from Eq. (1).

However, M is not invertible in general. A general regularized solution for the unfolding / quantification problem, with a regularization strength $\tau \geq 0$, is

$$\vec{p}^{\,*} \;=\; \underset{\vec{p}\geq 0 \text{ s.t. } \mathbb{1}^\top \vec{p}=1}{\arg\min} \; \mathcal{L}(\vec{p}\,;\, \vec{q}, M) + \tau \cdot r(\vec{p}), \tag{4}$$

where the loss function $\mathcal{L} : \mathbb{R}^C \to \mathbb{R}$ and the regularization function $r : \mathbb{R}^C \to \mathbb{R}$ are still to be defined for each particular unfolding / quantification method. The constraints in Eq. (4) ensure that $\vec{p}^{\,*}$ represents a valid probability density. Our framework extends the one by Firat [Fi16] with regularization functions $r(\vec{p})$.

Adhering to this framework are the most important unfolding algorithms, namely

Regularized Unfolding (RUN) [Bl02; Bl85] RUN models the likelihood of solutions in terms of Poisson-distributed counts. Namely, we observe a histogram of counts $\bar{q} = N \cdot \vec{q} \in \mathbb{N}^F$, each element of which is modelled as being Poisson-distributed with the rate $\lambda_i = [M\bar{p}]_i$. This modelling gives rise to the negative log-likelihood function

$$\mathcal{L}^{\text{RUN}}(\vec{p}\,;\, \vec{q}, M) \;=\; \sum_{i=1}^{F} [M\bar{p}]_i - \bar{q}_i \ln [M\bar{p}]_i, \tag{5}$$

which RUN minimizes.

To ensure smooth solutions, RUN employs Tikhonov regularization. The Tikhonov matrix $T \in \mathbb{R}^{C\times C}$ is defined such that

$$r^{\text{RUN}}(\vec{p}) \;=\; \frac{1}{2}\,(T\vec{p}\,)^2 \;=\; \frac{1}{2} \sum_{i=2}^{C-1} \left([\vec{p}]_{i-1} - 2[\vec{p}]_i + [\vec{p}]_{i+1}\right)^2 . \tag{6}$$

Unfolding via Singular Value Decomposition (SVD) [HK96] The SVD-based unfolding employs the regularizer from Eq. (6) with a least squares loss

$$\mathcal{L}^{\text{SVD}}(\vec{p}\,;\, \vec{q}, M) \;=\; \left\| \frac{\vec{q} - M\vec{p}}{\vec{w}} \right\|_2^2, \tag{7}$$

which is weighted by a vector $\vec{w} \in \mathbb{R}^F$. For instance, a Poisson model can be realized through Poisson variances $\vec{w} = \sqrt{\vec{q}}$.

Iterative Bayesian Unfolding (IBU) [DA10; DA95] IBU revolves around an expectation maximization approach. Starting from a prior $\vec{p}^{\,(0)}$, it repeatedly updates the estimate $\vec{p}^{\,(k)}$ according to Bayes' theorem

$$[\vec{p}^{\,(k)}]_i \;=\; \sum_{j=1}^{F} \frac{[M]_{ij}[\vec{p}^{\,(k-1)}]_i}{\sum_{i'=1}^{C}[M]_{i'j}[\vec{p}^{\,(k-1)}]_{i'}} [\vec{q}]_j. \tag{8}$$

IBU implements regularization in two ways. First, through early stopping in combination with a smooth prior. For instance, starting from $\vec{p}^{\,(0)} = \frac{1}{C}$ and stopping before Eq. (8) converges will maintain the smoothness of $\vec{p}^{\,(0)}$ to some degree. Second, the intermediate estimates $\vec{p}^{\,(k)}$ are smoothed with a low-order polynomial.

The above algorithms do not specify the feature transformation $f(\vec{x}) \in \{1, \ldots, F\}$ through which $\vec{q}$ and M are defined; they solely focus on the estimation of $\vec{p}$ from any given $\vec{q}$ and M. In this sense, these algorithms are open to any feature transformation. To this end, physicists have proposed

- to bin a single feature that is well correlated with the target quantity [Bl85],
- to cluster the features in order to map instances to cluster indices [Bu18],
- or to optimally partition the feature space by means of decision trees [Bö17]

in order to obtain histograms $\vec{q}$ which represent the data sample.

3 Quantification

In the following, we show that several algorithms from the quantification literature are indeed instances of our unified framework. A summary of these findings is displayed in Tab. 1. We prove the correctness of our unifying notation in the Appendix.

Tab. 1: Algorithms for unfolding and quantification within the framework of Eq. (4).

	loss function $\mathcal{L}$	regularizer r	feature transformation f
RUN [Bl02; Bl85]	$\sum_{i=1}^{d} [M\bar{p}]_i - \bar{q}_i \ln [M\bar{p}]_i$	$\frac{1}{2}\,(T\vec{p}\,)^2$	not specified / any
SVD [HK96]	$\left\| \frac{\vec{q} - M\vec{p}}{\vec{w}} \right\|_2^2$	$\frac{1}{2}\,(T\vec{p}\,)^2$	not specified / any
IBU [DA10; DA95]	expectation maximization	smoothing	not specified / any
ACC [Fo08; HK10]	$\|\vec{q} - M\vec{p}\,\|_2^2$	none	$\delta_{\arg\max_i [h(\vec{x})]_i}$
PACC [Be10]	$\|\vec{q} - M\vec{p}\,\|_2^2$	none	$h(\vec{x})$
ReadMe [HK10]	$\|\vec{q} - M\vec{p}\,\|_2^2$	none	$\delta_{\vec{x}=(X_1,\ldots,X_{2^d})}$
HDx [GAA13]	$\frac{1}{d}\sum_{i=1}^{d} \mathrm{HD}_i(\vec{q},\, M\vec{p}\,)$	none	$(\delta_{b(\vec{x};1)}, \ldots, \delta_{b(\vec{x};d)})$
HDy [GAA13]	$\frac{1}{d}\sum_{i=1}^{d} \mathrm{HD}_i(\vec{q},\, M\vec{p}\,)$	none	$(\delta_{b(h(\vec{x});1)}, \ldots, \delta_{b(h(\vec{x});C)})$
CC [Fo08]	none (assume $M = \mathbb{I}$)	none	$\delta_{\arg\max_i [h(\vec{x})]_i}$
PCC [Be10]	none (assume $M = \mathbb{I}$)	none	$h(\vec{x})$

Namely, our framework from Eq. (4) accommodates the following algorithms:

Adjusted Classify and Count (ACC) [Fo08; HK10] Hopkins; King [HK10] present a method that extends the binary adjustment by Forman [Fo08] to multi-class settings. Their extension represents a data sample as the counts of classification outcomes $\arg\max_i [h(\vec{x})]_i$, as returned by a multi-class classifier $h : \mathcal{X} \to \mathbb{R}^C$. In this case, M is simply the normalized confusion matrix of h, as estimated on hold-out training data. Hopkins; King [HK10] propose to solve Eq. (3) via constrained least squares regression, hence

$$\mathcal{L}^{\text{ACC}}(\vec{p}\,;\, \vec{q}, M) \;=\; \|\vec{q} - M\vec{p}\,\|_2^2 \tag{9}$$

and no regularization is employed.

Others have proposed to solve Eq. (3) through a matrix inversion, as $\vec{p}^{\,\text{inversion}} = M^{-1}\vec{q}$ [Mc92; VO01]. However, there is no guarantee that M is indeed invertible. Therefore, $\vec{p}^{\,\text{inversion}}$ might be undefined and the method by Hopkins; King [HK10] should be the prefered multi-class version of ACC.

Probabilistic ACC (PACC) [Be10] PACC employs the same adjustment as ACC, hence the same loss. However, PACC averages soft classifications $h(\vec{x}) \in \mathbb{R}^C$ instead of counting the crisp outcomes $\arg\max_i [h(\vec{x})]_i$.

ReadMe [HK10] Building on the multi-class version of ACC, ReadMe employs the loss function from Eq. (9). However, ReadMe transforms the features in a unique way that is motivated in text mining. In this application area, instances $\vec{x}$ are often represented as bags of words, i.e. by sparse indicator vectors $\{0,1\}^d$ for a vocabulary of size d. In ReadMe, $\vec{q}$ is a histogram over all 2^d possible incarnations X_i of these indicator vectors, i.e.

$$f^{\text{ReadMe}}(\vec{x}) \;=\; \delta_{\vec{x}=(X_1,\ldots,X_{2^d})}, \quad \text{where } [\delta_{a=(a_1,\ldots,a_n)}]_i \;=\; \begin{cases} 1 & \text{if } a = a_i, \\ 0 & \text{otherwise} \end{cases}. \tag{10}$$

Since such a representation is only feasible with small d, ReadMe produces multiple estimates, each of which employs a different and small random selection of words. Finally, all of these estimates are averaged.

HDx [GAA13] In this method, each feature is separately binned and a data sample is represented as a concatenation of all feature-wise histograms

$$f(\vec{x}) \;=\; (\delta_{b(\vec{x};1)}, \ldots, \delta_{b(\vec{x};d)}), \tag{11}$$

where $b(\vec{x}; i)$ is a binning function which maps the feature value $[\vec{x}]_i$ to the corresponding bin index $\{1, \ldots, B_i\}$.

The loss is measured as the average of feature-wise Hellinger distances,

$$\mathcal{L}(\vec{p}\,;\, M, \vec{q}) \;=\; \frac{1}{d}\sum_{i=1}^{d} \mathrm{HD}_i(\vec{q},\, M\vec{p}), \tag{12}$$

$$\text{where} \quad \mathrm{HD}_i(\vec{q},\, M\vec{p}) \;=\; \sqrt{\sum_{j=1+\sum_{k=1}^{i-1} B_k}^{\sum_{k=1}^{i} B_k} \left(\sqrt{[\vec{q}]_j} - \sqrt{[M\vec{p}]_j}\right)^2}. \tag{13}$$

HDy [GAA13] Originally, HDy has been proposed for binary quantification only. However, we can easily extend the method to the multi-class setting. In this setting, HDy replaces the separated binning of features $b(\vec{x}, i)$ in HDx with a separated binning of class-wise classifier outputs $b(h(\vec{x}), i)$. All other aspects of HDx are maintained.

(Probabilistic) Classify and Count (PCC/CC) [Be10; Fo08] We also conceive these non-adjusted methods, which simply return $\vec{q}$ as their estimates for $\vec{p}$, as instances of our framework. Strictly speaking, CC and PCC do not require the minimization of a loss function. More loosely speaking, however, their disregard of M can be understood as the assumption of a perfect classifier, so that $M = \mathbb{I}$ is the identity matrix. Under this assumption, the least squares loss from Eq. (9) leads to the estimate $\vec{p}^{\,\mathrm{CC}} = \vec{q}$ and we can understand this estimate as an instance of Eq. (4).

Regarding $f(\vec{x})$, CC employs the feature transformation of ACC and PCC employs the feature transformation of PACC.

4 Conclusion and Outlook

We have presented the unfolding algorithms RUN, SVD, and IBU and the quantification algorithms ACC, PACC, ReadMe, HDx, HDy, CC, and PCC within a common framework. These algorithms differ in terms of the loss functions, the regularizers, and the feature transformations they employ.

Our unification demonstrates the similarity between the problems that are approached in unfolding and in quantification literature. Due to this similarity, we conceive adaptations of quantification algorithms to physics problems as a valuable endeavor for future work. Likewise, we suggest to adapt unfolding algorithms to problems outside of physics.

References

[Be10] Bella, A.; Ferri, C.; Hernández-Orallo, J.; Ramírez-Quintana, M. J.: Quantification via Probability Estimators. In: Int. Conf. on Data Mining. IEEE, pp. 737–742, 2010.

[Bl02] Blobel, V.: An unfolding method for high energy physics experiments. In: Adv. Stat. Tech. in Part. Phys. Pp. 258–267, 2002.

[Bl85] Blobel, V.: Unfolding methods in high-energy physics experiments, tech. rep., CERN, 1985.

[Bö17] Börner, M.; Hoinka, T.; Meier, M.; Menne, T.; Rhode, W.; Morik, K.: Measurement/Simulation Mismatches and Multivariate Data Discretization in the Machine Learning Era. In: Astron. Data Anal. Softw. and Syst. ASP Conference Series, pp. 431–434, 2017.

[Bu18] Bunse, M.; Piatkowski, N.; Morik, K.; Ruhe, T.; Rhode, W.: Unification of deconvolution algorithms for Cherenkov astronomy. In: Int. Conf. on Data Sci. and Adv. Anal. IEEE, pp. 21–30, 2018.

[DA10] D'Agostini, G.: Improved iterative Bayesian unfolding. CoRR abs/1010.0632/, 2010, arXiv: 1010.0632.

[DA95] D'Agostini, G.: A multidimensional unfolding method based on Bayes' theorem. Nucl. Instr. and Meth. in Phys. Res. Sect. A 362/2-3, pp. 487–498, 1995.

[Fi16] Firat, A.: Unified Framework for Quantification. CoRR abs/1606.00868/, 2016, arXiv: 1606.00868.

[Fo08] Forman, G.: Quantifying counts and costs via classification. Data Mining and Knowl. Discov. 17/2, pp. 164–206, 2008.

[GAA13] González-Castro, V.; Alaíz-Rodríguez, R.; Alegre, E.: Class distribution estimation based on the Hellinger distance. Inf. Sci. 218/, pp. 146–164, 2013.

[Go17] González, P.; Castaño, A.; Chawla, N. V.; del Coz, J. J.: A Review on Quantification Learning. ACM Comput. Surv. 50/5, 74:1–74:40, 2017.

[GS16] Gao, W.; Sebastiani, F.: From classification to quantification in tweet sentiment analysis. Soc. Netw. Anal. and Mining 6/19, pp. 1–22, 2016.

[HK10] Hopkins, D. J.; King, G.: A method of automated nonparametric content analysis for social science. Amer. J. of Polit. Sci. 54/1, pp. 229–247, 2010.

[HK96] Hoecker, A.; Kartvelishvili, V.: SVD approach to data unfolding. Nucl. Instr. and Meth. in Phys. Res. Sect. A 372/3, pp. 469–481, 1996.

[Mc92] McLachlan, G. J.: Discriminant analysis and statistical pattern recognition. Wiley, 1992, ISBN: 0-471-69115-1.

[VO01] Vucetic, S.; Obradovic, Z.: Classification on data with biased class distribution. In: Europ. Conf. on Mach. Learn. Springer, pp. 527–538, 2001.

Acknowledgement: This work has been supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Data Analysis", project C3. https://sfb876.tu-dortmund.de.

A Proofs

We now detail the mapping from the original algorithms to our unified notation, to formally prove that our framework is consistent with the original proposals.

Regularized Unfolding (RUN) The loss function we present in Eq. (5) is a verbatim statement by Blobel [Bl85, Eqs. (2.29), and (2.26)]. The original algorithm treats the elements of $\vec{p}$ as B-spline coefficients; however, a more recent version by the same author [Bl02] employs histograms, which are consistent with our Eq. (2). Due to this change "the second derivative in bin j is proportional to $x_{j-1} - 2x_j + x_{j+1}$" [Bl02], where $x_i = [\vec{p}]_i$. This derivative defines the regularization term from Eq. (6). □

Unfolding via Singular Value Decomposition (SVD) The loss function we present in Eq. (7) and the regularization term from Eq. (6) are verbatim statements by Hoecker; Kartvelishvili [HK96, Eqs. (29), (37), and (38)]. □

Iterative Bayesian Unfolding (IBU) D'Agostini [DA95, Eqs. (3), and (4)] estimates $[\vec{p}^{\,(k)}]_i$ as

$$\frac{1}{\epsilon_i} \sum_{j=1}^{n_E} n(E_j) \cdot \frac{P(E_j \mid C_i) \cdot P_0(C_i)}{\sum_{l=1}^{n_C} P(E_j \mid C_l) \cdot P_0(C_l)},$$

where we identify our notation as $F = n_E$, $C = n_C$, $M_{ij} = P(E_j \mid C_i)$, and $[\vec{p}^{\,(k-1)}]_i = P_0(C_i)$. In the original algorithm, $n(E_j) \in \mathbb{N}$ is the count observed in the j-th bin, i.e. $n(E_j) = N \cdot [\vec{q}]_j$. Moreover, $\epsilon_i > 0$ is an acceptance factor, which models the probability that an existing instance of class i is indeed part of the sample—and not hidden due to measurement complications. Setting $\epsilon_i = N$, we obtain $[\vec{q}]_j = \frac{n(E_j)}{\epsilon_i}$, which is consistent with our Eq. (8).

For regularization, D'Agostini [DA95] proposes to "smooth the results of the unfolding before feeding them in the next step", for instance "by a polynomial fit of 3rd degree" or by another low-order polynomial. □

Adjusted Classify and Count (ACC) Hopkins; King [HK10, Eq. (4)] marginalize over the true labels $D \in \{1, \ldots, J\}$ to yield the distribution of class predictions $\widehat{D}$,

$$P(\widehat{D} = j) \;=\; \sum_{j'=1}^{J} P(\widehat{D} = j \mid D = j') P(D = j).$$

The authors note that "this expression represents a set of J equations […] that can be solved for the J elements in $P(D)$". Accordingly, we identify our notation as $\vec{p} = P(D)$, $\vec{q} = P(\widehat{D})$, and $M = P(\widehat{D} \mid D)$ in their presentation. To solve this set of equations, the authors propose a "standard constrained least squares to ensure that elements of $P(D)$ are each in [0,1] and collectively sum up to 1". This proposal defines the least squares loss from Eq. (9) and matches our constraints in Eq. (4). □

Note that Hopkins and King have developed their method independently of Forman's binary ACC. However, the basis of their work is precisely the adjustment by Forman [Fo08, Eq. (1)], as can be seen in Hopkins; King [HK10, Eq. (3)]. Therefore, we call their method "multi-class ACC".

The other multi-class extension of ACC, $\vec{p}^{\,\text{inversion}}$, is presented in McLachlan [Mc92, Eq. (2.3.4)] and in Vucetic; Obradovic [VO01, Eq. (3)].

Probabilistic ACC (PACC) The essential proposal by Bella et al. [Be10] is to replace hard classifications $\arg\max_i [h(\vec{x})]_i$ with probabilistic ones $h(\vec{x}) \in \mathbb{R}^C$; their adjustment is the same as in binary ACC. Taking this proposal to multi-class ACC [HK10], we obtain multi-class PACC with the loss from Eq. (9). □

ReadMe Building on their multi-class design of ACC, Hopkins; King [HK10, Eq. (6)] set up a matrix equation $P(\mathbf{S}) = P(\mathbf{S} \mid D)P(D)$, which maps to our notation as $\vec{q} = P(\mathbf{S}) \in \mathbb{R}^{2^d}$, $M = P(\mathbf{S} \mid D) \in \mathbb{R}^{2^d \times C}$, and $\vec{p} = P(D) \in \mathbb{R}^C$. The authors note that "$P(\mathbf{S})$ is the probability of each of the 2^K possible word stem profiles" with $K = d$ being the number of word stems. To estimate this probability, "we merely compute the proportion of documents observed with each pattern of word profiles". This computation leads to a histogram

$$\vec{q} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\vec{x}_i = (X_1, \ldots, X_{2^d})},$$

which is consistent with our Eqs. (2) and (10). □

HDx González-Castro et al. [GAA13, Eq. (9)] minimize the average of feature-wise Hellinger distances, as we have stated in Eq. (12). They present the distance with respect to a single feature j, binned into b bins, as

$$\sqrt{\sum_{i=1}^{b} \left(\sqrt{\frac{|V_{j,i}|}{|V|}} - \sqrt{\frac{|U_{j,i}|}{|U|}} \right)^2},$$

where $|U|$ is the total number of instances and $|U_{j,i}|$ is the number of instances whose feature j is mapped to the i-th bin [GAA13, Eq. (10)]. $|V|$ and $|V_{j,i}|$ are the numbers of instances that are to be expected under class prevalences $\vec{p}$, hence

$$\frac{|V_{j,i}|}{|V|} = [M\vec{p}]_{i + \sum_{k=1}^{j-1} B_k},$$

where $\sum_{k=1}^{j-1} B_k$ is the offset of the histogram of feature j within our concatenation of feature-wise histograms. Using the product $M\vec{p}$ at this point is consistent with the binary conception of González-Castro et al. [GAA13, Eq. (12)]. □

HDy The original HDy [GAA13, Eqs. (13) and (14)] only addresses binary quantification. For this case, however, the only change with respect to HDx is that HDy employs soft classifier outputs $h(\vec{x})$ instead of features $\vec{x}$. A straightforward extension to the multi-class setting is therefore to bin the class-wise outputs $[h(\vec{x})]_i$ separately, as HDx does in case of features and as we propose in our presentation of HDy. □

(Probabilistic) Classify and Count (PCC/CC) Let $M = \mathbb{I}$. Recognize that the global minimum of the least squares loss,

$$\min_{\vec{p}} \|\vec{q} - M\vec{p}\|_2^2 = 0,$$

is now attained if and only if $\vec{p} = \vec{q}$. Therefore, under the assumption $M = \mathbb{I}$, the unique minimizer of the least squares loss is $\vec{q}$. In this sense, PCC and CC are proper instances of our framework. □

Radio Galaxy Classification with wGAN-Supported Augmentation

Janis Kummer[1,2], Lennart Rustige[1,3], Florian Griese[1,4], Kerstin Borras[3,5], Marcus Brüggen[2], Patrick L. S. Connor[1,6], Frank Gaede[3], Gregor Kasieczka[6], Peter Schleper[6]

Abstract: Novel techniques are indispensable to process the flood of data from the new generation of radio telescopes. In particular, the classification of astronomical sources in images is challenging. Morphological classification of radio galaxies could be automated with deep learning models that require large sets of labelled training data. Here, we demonstrate the use of generative models, specifically Wasserstein GANs (wGAN), to generate artificial data for different classes of radio galaxies. Subsequently, we augment the training data with images from our wGAN. We find that a simple fully-connected neural network for classification can be improved significantly by including generated images into the training set.

Keywords: Radio galaxy classification, Generative models, GANplyfication

1 Introduction

The new generation of radio telescopes (e.g. LOFAR, MeerKAT and in the future the SKA [Ca04; JM16; va13]) will produce enormous amounts of data and the improved sensitivity of the instruments leads to a much higher source density within these data sets. As a result, a new level of automation for processing the data and for classifying sources is needed. Morphological classification of radio sources is crucial for answering a range of fundamental astrophysical questions, such as the origin of cosmic magnetism. One promising approach for morphological classification relies on the use of deep classifiers trained on well-understood data sets. However, the existing amount of data with morphological labels is limited, as they are typically extracted from catalogues created and curated manually by experts. Small data

[1] Center for Data and Computing in Natural Sciences (CDCS), Notkestrasse 9, D-22607 Hamburg, Germany
[2] Universität Hamburg, Hamburger Sternwarte, Gojenbergsweg 112, D-21029 Hamburg, Germany
janis.kummer@uni-hamburg.de
[3] Deutsches Elektronen-Synchrotron DESY, Notkestrasse 85, D-22607 Hamburg, Germany
lennart.rustige@desy.de
[4] Section for Biomedical Imaging, University Medical Center Hamburg-Eppendorf, D-20246 Hamburg, Germany
Institute for Biomedical Imaging, Hamburg University of Technology, D-21073 Hamburg, Germany
florian.griese@tuhh.de
[5] RWTH Aachen University, Templergraben 55, D-52062 Aachen, Germany
[6] Universität Hamburg, Institut für Experimentalphysik, Luruper Chaussee 149, D-22761 Hamburg, Germany

doi:10.18420/inf2022_38

sets used in the training of deep learning models for galaxy classification can be enlarged by data augmentation, e.g. by applying random rotations and reflections to the images (classical augmentation). In this work, we investigate a novel application of generative models to enhance the available training sets. For this augmentation technique, multiple neural networks are combined to learn the underlying distribution of a data set. This allows the creation of new data by sampling from the learnt distribution. In this study we investigate whether radio galaxy classifiers can be improved when trained on such enhanced data sets containing generated data. For similar approaches from different fields see for example [Fr18; Zh17].

As radio sources exhibit a large variety of structures, we consider a four-class classification problem (as in [ATV18; Sa21]), including bent-tail and compact sources in addition to the classes FRI and FRII of Fanaroff; Riley [FR74]. For the class FRI, the maximum of the radio emission is situated close to the centre of the source. The maxima of the radio emission are located at the edges of the jets for FRII sources. Unresolved and point sources are contained in the Compact class. The Bent class consists of sources for which the angle between the jets differs significantly from 180 degrees. The two subtypes narrow-angle tail (NAT) and wide-angle tail (WAT) are further discriminated by the angle. For an illustration of the considered classes (FRI, FRII, Compact and Bent) see Fig. 1.

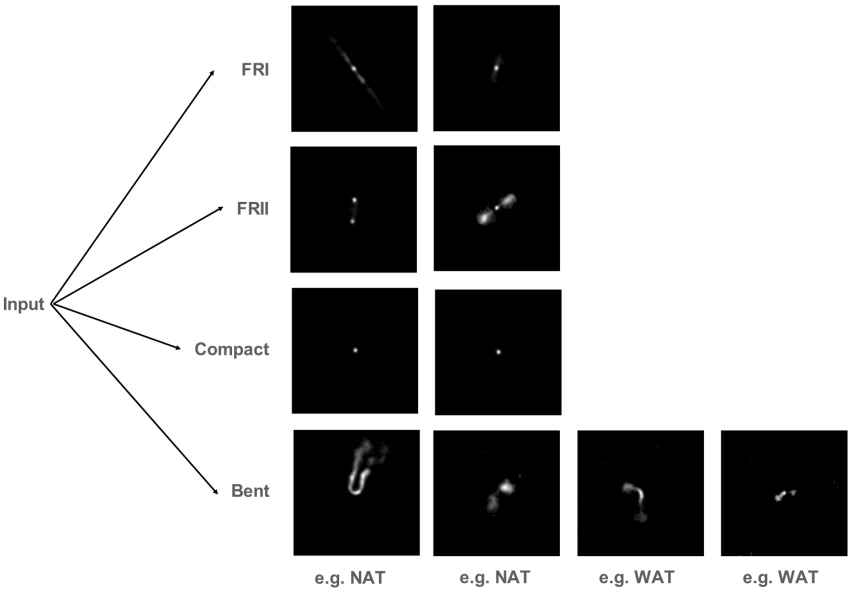


Fig. 1: Examples of the classes FRI, FRII, Compact and Bent.

Aniyan; Thorat [AT17], Alhassan et al. [ATV18], Samudre et al. [Sa21], and Tang et al. [TSL19] use convolutional neural networks (CNNs) trained on data from the FIRST survey [BWH95] for the classification of radio galaxies.

2 Data

We combine different catalogues that characterise radio sources from the FIRST survey to create a data set of radio galaxy images with morphological labels [BCM17; CMB17a; CMB17b; GBW10; GW08; MB17; Pr11]. By explicitly comparing the coordinates of the sources, 300 duplicates were found and removed. Moreover, we exclude 147 sources from the data set that appear in different catalogues with different labels. The resulting data sets are presented in Table 1. We adopt the preprocessing from [AT17] after cropping the images to the input size of our generative network (128 x 128 pixels). In particular, we set all pixel values below three times the local RMS noise to the value of this threshold. Subsequently, the pixel values are rescaled to the range between -1 and 1 to represent floating point greyscale images. We apply classical augmentation to each image during training.

	FRI	FRII	Compact	Bent	Total
train	395	824	291	248	1758
validation	50	50	50	50	200
test	50	50	50	50	200
total	495	924	391	348	2158

Tab. 1: Number of sources per class in the data sets.

3 Wasserstein GAN

Generative adversarial networks (GANs) [Go14; Sa16] are able to learn a representation of the underlying statistical distributions of sets of images. Sampling from those representations may provide additional data points for subsequent treatments [Bu21a; Bu22].

For this project, we employ a variant of the standard GAN setup called Wasserstein GAN (wGAN), which uses the Wasserstein-1 metric, often referred to as the Earth Mover's distance, as main term in the loss function [ACB17]. A direct advantage of this setup is the correlation between image quality and the value of the loss function, transforming the discriminant of a standard GAN into a critic. Additionally, training of wGANs is often more stable and more likely to converge than standard GAN setups. To approximate the Wasserstein-1 metric by use of a critic network, it has to be ensured that the 1-Lipschitz constraint is fulfilled. This is achieved by applying a gradient penalty term to the loss function as in [Gu17].

Since different morphologies of radio galaxies result in very different images, it seems reasonable to condition the networks with the class label. For this setup, this is achieved for the generator by applying a 2D transposed convolution operator on a matrix of image dimensions filled with the class label. The transpose-convoluted layer is then concatenated to the first transpose-convoluted layer of the noise tensor. Batch normalisation in 2D and ReLU activation functions are used. The concatenated tensor is then passed through five additional 2D transposed convolutions, where no normalisation or activation is applied

after the last layer. Instead, the individual pixel values are simply clipped to [-1,1] for easy conversion to greyscale. The critic is built analogously, but uses 2D convolutional layers resulting in a single output node representing the critic score for image quality. Here, layer normalisation and leaky ReLU functions are used except for the last layer. A schematic of the wGAN setup can be found in Fig. 2.

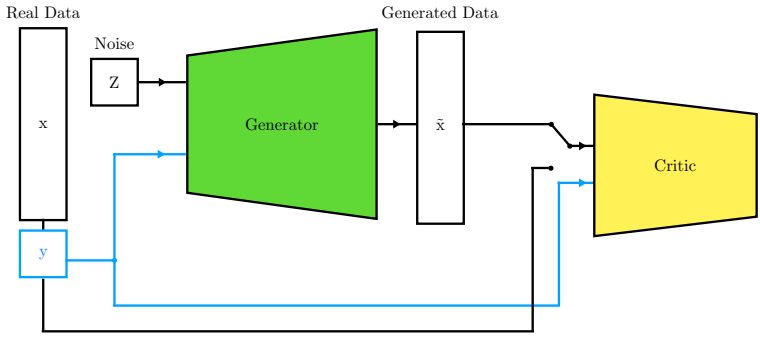


Fig. 2: Schematic of the wGAN architecture, where x denotes real images and y the class label. The noise vector Z is the input of the generator and the output of the generator $\tilde{x}$ represents generated images. The critc receives either real or generated images as input.

4 Generated Images

4.1 Training

To evaluate statistical fluctuations, ten statistically independent wGAN training runs are launched solely on the training set. The training is performed with a single NVIDIA A100 GPU provided by the Maxwell cluster at DESY. The training is done for 40,000 generator iterations, while the critic is trained five times per generator iteration. A batch size of 400 is chosen and the full training takes roughly 7 hours to complete. The generator and critic weights are saved every 250 iterations, allowing to scan for the best training state later on.

4.2 Image Quality

In order to determine the quality of images and thus to find the best performing training iteration, a set of distributions is defined to compare generated images to the validation data set. This includes normalised histograms of pixel intensities, the number of pixels with an intensity greater than zero and of the sum of intensities. These histograms are compared for each class and the relative mean absolute error (RMAE) between the generated set of 10,000 images and the validation set is computed. The RMAE for the different histograms are

summed up to yield a single figure-of-merit (FOM) per class, where the training iteration with the lowest FOM value is used in the following. As an example, the histogram of pixel intensities for FRI is depicted in Fig. 3, where the generated histogram is shown in blue and the validation set in orange.

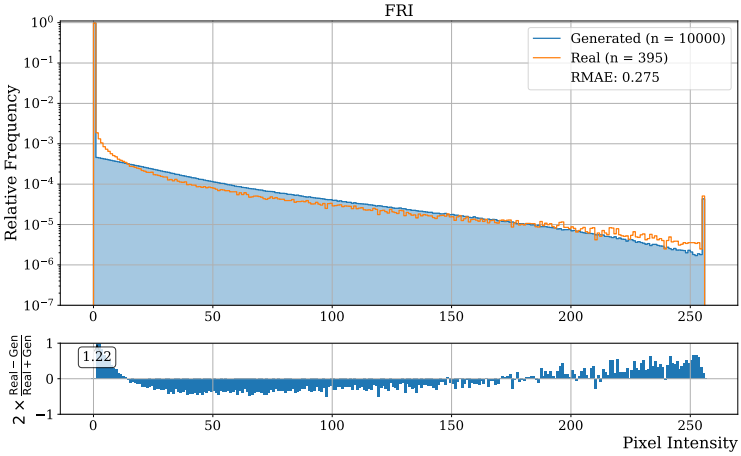


Fig. 3: Histogram of pixel intensities of FRI sources comparing batches of real (orange) and generated images (blue) for the training epoch with the lowest combined RMAE. The per-bin relative error is shown in the bottom panel.

For a direct visual comparison of image quality, a set of 5,000 images per class is generated and aligned using principle component analysis. Subsequently, the pixel-by-pixel difference is computed for all possible pairs of real and generated images. The closest pairs with a minimal activity for each class are shown in Fig. 4.

5 Using wGAN-supported Augmentation

As described above, classical augmentation relies on rotation and reflection of training images, adding no statistical information on the morphology. This type of augmentation is, thus, mainly a way to help a classifier understand the rotational invariant nature of radio galaxy images. Nowadays, this could be achieved by directly using equivariant models that incorporate such symmetry constraints inherently and show potential to improve classifier training without relying exclusively on augmentation [Bo21].

Here, we evaluate the augmentation achieved by using a combined data set of real and generated images (i.e. wGAN-supported augmentation) by comparing it to using the classically augmented (real-only) data set. The combined data set is created before starting the classifier training with a fixed ratio between the number of generated i_g and real images i_r, denoted $\lambda = i_g/i_r$. We study the range $\lambda = 1, ..., 4$ and the class distribution of the

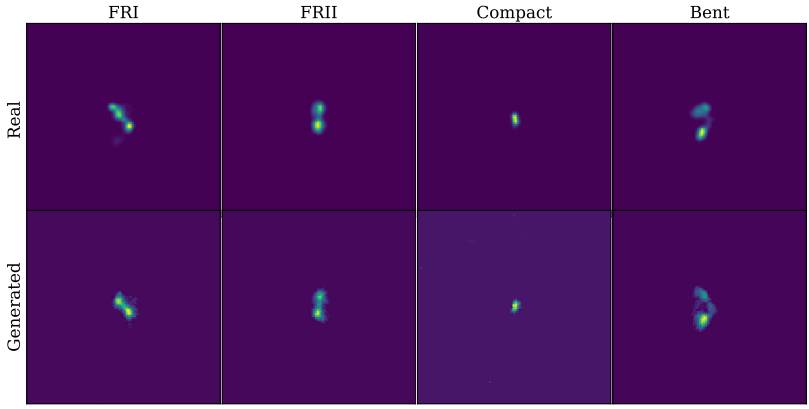


Fig. 4: Closest matching pairs of generated and real images.

generated images is designed such that the combined data set is balanced. For the real-only training runs, class weights are introduced to the loss function to account for the class imbalance in the data set.

We train a simple classifier, namely a fully-connected neural network with three hidden layers and four output nodes. The classifier is trained for 12,500 epochs with a batch size of 100 and a learning rate of $5 \cdot 10^{-4}$. Training takes on average 3.5 seconds per epoch on an NVIDIA V100 GPU for real-only runs. This setup is not optimised to reach maximal classification accuracies as the goal of this study is only to compare classical with wGAN-supported augmentation.

We observe a significant improvement when increasing the number of generated images in the combined data set. As shown in Fig. 5 the real-only training run enters into over-training (i.e. the value of the loss function increases, while the training accuracy deviates further from the validation accuracy) much earlier than training runs using a combined training data set. The start of the over-trained period depends on λ, where we observe that adding more generated images to the data set shifts the start to later training iterations, while the validation loss score continues to decrease and the validation accuracy score (not shown) increases.

In order to compare the overall performance between different training setups, the multi-class Brier score [Br50] is used to determine the best training iteration for each of the ten statistically independent training runs for our five different training setups. The Brier score is essentially the mean squared error of the predicted probabilities of a classifier for all classes. This has the advantage that also the certainty of the classifier's decision is considered, which winner-takes-all FOMs such as accuracy do not take into account. The performance of the best models (i.e. models with minimal Brier score) is evaluated on

an independent test set that contains real data only using commonly applied metrics such as accuracy, precision, recall, and F_1 score. The F_1 score is the harmonic mean between precision and recall. The results for the F_1 score are depicted in Fig. 6, indicating that training runs using the combined data set with $\lambda = 4$ achieve an F_1 score that is (23 ± 2) % higher than for the real-only training runs.

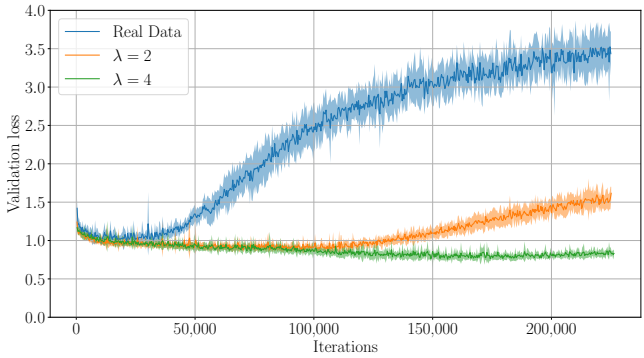


Fig. 5: Validation loss during training of the classifier for different values of λ. The solid line represents the mean of ten independent training runs and the shaded area represents the symmetrised standard deviation.

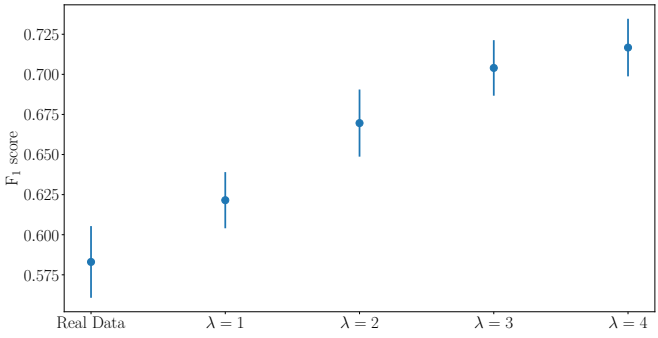


Fig. 6: Performance of the best model on an independent test set. We show the mean of ten independent training runs, where the uncertainty is given as the symmetrised standard deviation.

6 Discussion

Previous approaches employing generative models to simulate images of radio galaxies are based on variational autoencoders (VAE) [Ba21; Ma18; Ma19]. Therefore our GAN-based approach is novel in the field of radio astronomy. In this work, we demonstrate that we

are able to generate highly realistic, artificial images of radio sources. The quality of the generated images is evaluated in different ways and with different metrics. After visual inspection of a large number of generated images, our generated images do not suffer from issues known from other state-of-the-art simulations based on generative networks. In summary, we find:

- The resolution of the generated images is as good as the resolution of the images we train on.
- The generated images have a similar noise level as the preprocessed training examples.
- We do not observe any pseudo-textures or pseudo-structures in the generated images.

The images generated with the wGAN can be put to use for applications such as classifier training. As a result, our contribution represents a major improvement for simulations of radio galaxies based on generative networks. Butter et al. [Bu21b] showed for toy models that an amplification of training statistics with generative models is possible. Our work represents an example of this GANplyfication with real astrophysical data. For a recent particle physics application see [Bi22].

The presented method allows to amplify the performance of a simple classification model substantially by an increase in the F_1 score of 23 % for a ratio of $\lambda = 4$ between generated and real images and can be particularly useful for applications with unbalanced data sets. The application of wGAN-supported augmentation to more complex classifiers such as CNNs is left for future work.

Acknowledgements

This work was supported by UHH, DESY, TUHH and HamburgX grant LFF-HHX-03 to the Center for Data and Computing in Natural Sciences (CDCS) from the Hamburg Ministry of Science, Research, Equalities and Districts. This project benefits greatly from the exchange with particle physicists with a vast experience in using generative models for calorimeter simulations and was supported in part through the Maxwell computational resources operated at DESY.

References

[ACB17] Arjovsky, M.; Chintala, S.; Bottou, L.: Wasserstein GAN, 2017, arXiv: 1701.07875.

[AT17] Aniyan, A. K.; Thorat, K.: Classifying Radio Galaxies with the Convolutional Neural Network. The Astrophysical Journal Supplement Series 230/2, p. 20, June 2017.

[ATV18] Alhassan, W.; Taylor, A. R.; Vaccari, M.: The FIRST Classifier: compact and extended radio galaxy classification using deep Convolutional Neural Networks. MNRAS 480/2, pp. 2085–2093, July 2018.

[Ba21] Bastien, D. J.; Scaife, A. M. M.; Tang, H.; Bowles, M.; Porter, F.: Structured variational inference for simulating populations of radio galaxies. MNRAS 503/3, pp. 3351–3370, Mar. 2021.

[BCM17] Baldi, R. D.; Capetti, A.; Massaro, F.: FR0CAT: a FIRST catalog of FR 0 radio galaxies. A&A 609/, A1, Dec. 2017.

[Bi22] Bieringer, S.; Butter, A.; Diefenbacher, S.; Eren, E.; Gaede, F.; Hundhausen, D.; Kasieczka, G.; Nachman, B.; Plehn, T.; Trabs, M.: Calomplification – The Power of Generative Calorimeter Models, 2022, arXiv: 2202.07352.

[Bo21] Bowles, M.; Bromley, M.; Allen, M.; Scaife, A.: E(2) Equivariant Self-Attention for Radio Astronomy, 2021, arXiv: 2111.04742.

[Br50] Brier, G. W.: Verification of forecasts expressed in terms of probability. Monthly Weather Review 78/1, pp. 1–3, 1950.

[Bu21a] Buhmann, E.; Diefenbacher, S.; Eren, E.; Gaede, F.; Kasieczka, G.; Korol, A.; Krüger, K.: Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed. Comput Softw Big Sci 5/13, May 2021.

[Bu21b] Butter, A.; Diefenbacher, S.; Kasieczka, G.; Nachman, B.; Plehn, T.: GANplifying event samples. SciPost Physics 10/6, June 2021.

[Bu22] Buhmann, E.; Diefenbacher, S.; Eren, E.; Gaede, F.; Hundhausen, D.; Kasieczka, G.; Korcari, W.; Krüger, K.; McKeown, P.; Rustige, L.: Hadrons, Better, Faster, Stronger. Mach. Learn.: Sci. Technol. 3/2, July 2022.

[BWH95] Becker, R. H.; White, R. L.; Helfand, D. J.: The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. The Astrophysical Journal 450/, p. 559, Sept. 1995.

[Ca04] Carilli, C.; Furlanetto, S.; Briggs, F.; Jarvis, M.; Rawlings, S.; Falcke, H.: Probing the dark ages with the Square Kilometer Array. New Astronomy Reviews 48/11-12, pp. 1029–1038, Dec. 2004.

[CMB17a] Capetti, A.; Massaro, F.; Baldi, R. D.: FRICAT: A FIRST catalog of FR I radio galaxies. A&A 598/, A49, Jan. 2017.

[CMB17b] Capetti, A.; Massaro, F.; Baldi, R. D.: FRIICAT: A FIRST catalog of FR II radio galaxies. A&A 601/, A81, May 2017.

[Fr18] Frid-Adar, M.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H.: Synthetic Data Augmentation using GAN for Improved Liver Lesion Classification, 2018, arXiv: 1801.02385.

[FR74] Fanaroff, B. L.; Riley, J. M.: The morphology of extragalactic radio sources of high and low luminosity. MNRAS 167/, 31P–36P, May 1974.

[GBW10] Gendre, M. A.; Best, P. N.; Wall, J. V.: The Combined NVSS-FIRST Galaxies (CoNFIG) sample - II. Comparison of space densities in the Fanaroff-Riley dichotomy. MNRAS 404/4, pp. 1719–1732, Mar. 2010.

[Go14] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y.: Generative Adversarial Networks, 2014, arXiv: 1406.2661.

[Gu17] Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.: Improved Training of Wasserstein GANs, 2017, arXiv: 1704.00028.

[GW08] Gendre, M. A.; Wall, J. V.: The Combined NVSS-FIRST Galaxies (CoNFIG) sample - I. Sample definition, classification and evolution. MNRAS 390/2, pp. 819–828, Sept. 2008.

[JM16] Jonas, J.; MeerKAT Team: The MeerKAT Radio Telescope. In: MeerKAT Science: On the Pathway to the SKA. 1, p. 1, Jan. 2016.

[Ma18] Ma, Z.; Zhu, J.; Li, W.; Xu, H.: Radio Galaxy Morphology Generation using Residual Convolutional Autoencoder and Gaussian Mixture Models. In: 2018 25th IEEE ICIP. Pp. 3044–3048, 2018.

[Ma19] Ma, Z.; Zhu, J.; Zhu, Y.; Xu, H.: Radio Galaxy Morphology Simulation via Residual Conditional Variational Autoencoder. In: 2019 15th CIS). Pp. 151–155, 2019.

[MB17] Miraghaei, H.; Best, P. N.: The nuclear properties and extended morphologies of powerful radio galaxies: the roles of host galaxy and environment. MNRAS 466/4, pp. 4346–4363, Jan. 2017.

[Pr11] Proctor, D. D.: Morphological Annotations For Groups In The FIRST Database. The Astrophysical Journal Supplement Series 194/2, p. 31, May 2011.

[Sa16] Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X.: Improved Techniques for Training GANs, 2016, arXiv: 1606.03498.

[Sa21] Samudre, A.; George, L. T.; Bansal, M.; Wadadekar, Y.: Data-Efficient Classification of Radio Galaxies. MNRAS 509/2, pp. 2269–2280, Oct. 2021.

[TSL19] Tang, H.; Scaife, A. M. M.; Leahy, J. P.: Transfer learning for radio galaxy classification. MNRAS 488/3, pp. 3358–3375, Sept. 2019.

[va13] van Haarlem, M. P. et al.: LOFAR: The LOw-Frequency ARray. A&A 556/, A2, Aug. 2013.

[Zh17] Zhu, X.; Liu, Y.; Qin, Z.; Li, J.: Data Augmentation in Emotion Classification Using Generative Adversarial Networks, 2017, arXiv: 1711.00648.

Towards probabilistic multiclass classification of gamma-ray sources

Dmitry Malyshev[1], Aakash Bhat[2]

Abstract:

Machine learning algorithms have been used to determine probabilistic classifications of unassociated sources. Often classification into two large classes, such as Galactic and extra-galactic, is considered. However, there are many more physical classes of sources (23 classes in the latest Fermi-LAT 4FGL-DR3 catalog [Ab22]). In this note we subdivide one of the large classes into two subclasses in view of a more general multi-class classification of gamma-ray sources. Each of the three large classes still encompasses several of the physical classes. We compare the performance of classifications into two and three classes. We calculate the receiver operating characteristic curves for two-class classification, where in case of three classes we sum the probabilities of the sub-classes in order to obtain the class probabilities for the two large classes. We also compare precision, recall, and reliability diagrams in the two- and three-class cases.

Keywords: Multiclass classification, random forest, neural networks, gamma-ray sources

1 Introduction

About one third of gamma-ray sources in Fermi Large Area Telescope (LAT) catalogs are unassociated [AAA+15; AAA+20; Ab22]. Machine learning methods trained with associated sources have been used to determine most likely classes of unassociated sources or, more generally, the probabilities for unassociated sources to belong to different classes [AAA+12; BM22; FKM21; LP17; Lu20; Mi16; Sa16; ZKZ21]. One of the main challenges in this approach is that the number of sources is relatively small: 6658 sources in the Fourth Fermi LAT source catalog data release 3 (4FGL-DR3) [Ab22], while the number of classes is relatively large: 23 classes in 4FGL-DR3 excluding unassociated sources and sources with "unknown" class, i.e., sources within 10° from the Galactic plane and associated to radio or X-ray sources of unknown nature [AAA+20]. In order to obtain sufficiently many sources for training and testing, the physical classes are combined in larger classes. In most of the analyses two classes are used, such as extra-galactic and Galactic sources, or active galactic nuclei (AGNs) and pulsars.

Subdivision of large classes into subclasses (which we refer to as "class subdivision" in the rest of the paper for brevity) was considered in the past, e.g., for sub-classification

[1] Erlangen Centre for Astroparticle Physics, Erwin-Rommel-Str. 1, Erlangen, Germany dmitry.malyshev@fau.de
[2] Erlangen Centre for Astroparticle Physics, Erwin-Rommel-Str. 1, Erlangen, Germany aakashbhat7@gmail.com

doi:10.18420/inf2022_39

of pulsars into young and millisecond ones [Lu20; Sa16] or for classification of Galactic sources into pulsars and other sources [ZKZ21]. The procedure of [Lu20; Sa16] included classification of all sources into AGNs and pulsars first, followed by dividing likely pulsars into young and millisecond pulsars by a separate ML training step. Apart from being a relatively complicated procedure, which is difficult to generalize beyond 3 classes, there is also a conceptual problem that the results depend on the choice of the threshold for the separation of pulsars and AGNs. Although such procedure makes sense for the search of the most likely candidates, the pulsar-like probabilities in sources classified as AGNs are not taken into account for the young vs millisecond pulsar classification, which prevents the interpretation of the final scores as probabilities and complicates the use of such catalogs for population studies including unassociated sources. In [ZKZ21], the 2-class classification included AGNs and pulsars, while the 3-class case included AGNs, pulsars, and other Galactic sources. The results of [ZKZ21] for the 3-class classification with two different ML algorithms lacked consistency. For instance, the random forest (RF) method gave 244 source candidates in the "other" class for 20 input features, while artificial neural networks (NNs) gave no candidates for the same input data. In [BM22] we have revisited the question of the 3-class classification into AGNs, pulsars, and other sources, and have shown that it is possible to have a consistent classification of sources into the three classes by several ML algorithms. Interestingly, we have also found that the reliability diagrams in the 3-class case and the systematic uncertainties related to the choice of the classification method were similar in the 3-class case compared to the 2-class case. In this work we extend the consistency checks of the 3-class classification using the latest 4FGL-DR3 catalog [Ab22]. In particular, we address the following question: can the 3-class classification also be used for the 2-class classification, where the 2-class probabilities are obtained by summing the probabilities of sub-classes in the 3-class case? For instance, would receiver operating characteristic (ROC) curve, precision, and recall for the 2-class case reduced from the 3-class case be similar, better, or worse compared to these characteristics derived in the 2-class classification?[3] For example, if the performance of the 2- and 3-class classifications for 2-class characteristics is similar, then the 3-class classification has a clear advantage over the 2-class classification since the 2-class probabilities can be recovered from the 3-class classification and this procedure does not degrade the 2-class classification performance. On the other hand, the 3-class classification provides additional information about probabilities of subclasses.

In this work we start with two large classes (AGN-like and Galactic sources) and test the effects on classification performance of dividing the Galactic sources into pulsars and other Galactic sources. One of the challenges of subdividing the classes is due to dataset imbalance - there are about seven times more AGN-like sources than Galactic ones. Dividing the Galactic sources into two classes leads to further imbalance of the datasets with imbalance factors on the order of 13 to 1 for AGN-like vs pulsars or 15 to 1 for AGN-like vs other Galactic sources. On the one hand the increased imbalance can lead to a decrease

[3] In this paper we use only associated sources to study the question of class subdivision. Application to probabilistic classification of unsassiciated sources will be reported in a follow up study.

in performance in the two-class classification, where the probabilities of the sub-classes of Galactic sources in the 3-class case are added to determine the overall probability for the source to be Galactic. On the other hand, the additional information contained in the sub-classes may lead to an improved two-class classification. For the classification we use RF and NN (with a single hidden layer) algorithms. We show that at least for the example considered in this paper, the 3-class classification does not lead to a decrease of performance compared to the 2-class classification, even if the subdivision of classes leads to an increase of imbalance among the classes.

The paper is organized as follows. In Section 2 we provide details about the training of the RF and NN algorithms including a description of classes and features. We also optimize the selection of meta-parameters by calculating the overall accuracy of predictions in the 2- and 3-class cases. In Section 3 we perform consistency checks of classification, when we sub-divide the Galactic class. In particular, we

1. Compare 2-class classification characteristics: (1) ROC curves and (2) precision and recall for Galactic sources as a function of thresholds on probability to belong to the Galactic class in cases when we subdivide and do not subdivide the Galactic class;
2. Calculate the precision and recall for the sub-classes of the Galactic class as a function of threshold on probability to belong to the corresponding sub-classes;
3. Calculate the reliability diagrams in the 2- and 3-class cases.

Section 4 contains conclusions and discussion.

2 Training

2.1 Definition of classes in the 4FGL-DR3 catalog

The 4FGL-DR3 catalog has 23 classes in addition to unassociated sources or sources with an unknown class [Ab22]. If a source is associated with a source at other frequencies and there is additional information, such as coincident flaring activity for AGNs or pulsed gamma-ray emission for pulsars, then the counterpart is said to be identified and the corresponding class is written in capital letters. Classes of sources based on positional associations only are written in lower-case letters. In this work we will treat both types of associations on equal footing and use lower-case class names for both identified and associated sources.

As the first step, we divide the sources into two large classes:

1. AGNs – 3814 sources (4FGL-DR3 class names: ’bll’, ’fsrq’, ’rdg’, ’agn’, ’ssrq’, ’css’, ’bcu’, ’nlsy1’, ’sey’)[4];

[4] For the full description of the classes, please see [Ab22].

2. Non AGN sources – 553 sources (4FGL-DR3 class names: 'gc', 'psr', 'msp', 'pwn', 'snr', 'spp', 'glc', 'sfr', 'hmb', 'lmb', 'bin', 'nov', 'sbg', 'gal'). Since the majority of non-AGN sources are inside the Milky way galaxy, we will denote this class as Galactic sources.

At the next step we divide the Galactic sources into two sub-classes

1. Pulsars – 292 sources (4FGL-DR3 class names: 'psr', 'msp');
2. Galactic non-pulsar sources – 261 sources (4FGL-DR3 class names: 'gc', 'pwn', 'snr', 'spp', 'glc', 'sfr', 'hmb', 'lmb', 'bin', 'nov', 'sbg', 'gal').

2.2 Algorithms

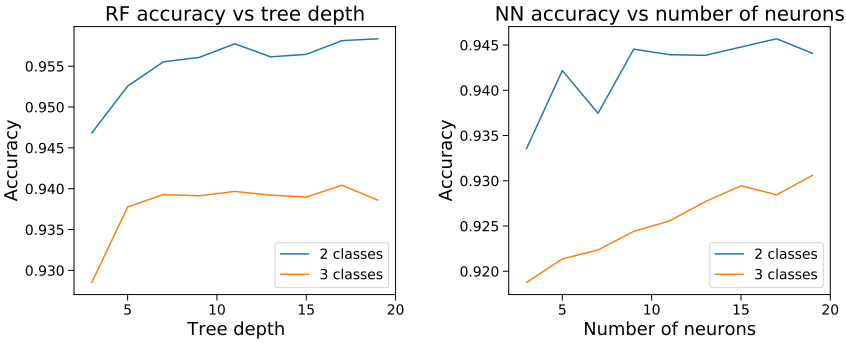


Fig. 1: RF accuracy as a function of maximal tree depth (left) and NN accuracy as a function of the number of neurons in the hidden layer (right).

In this work we use the RF algorithm with 50 trees and a NN with a single hidden layer implemented in scikit-learn package [Pe11]. The NN has the "tanh" activation function at the hidden layer and "softmax" activation function for the output layer. As input, we use the following 10 features from the 4FGL-DR3 catalog [Ab22] (some features are transformed in order to have a comparable range of values for the different features and in order to avoid discontinuity for the Galactic longitude): sin(GLAT), cos(GLON), sin(GLON), $\log_{10}$(Energy_Flux100), $\log_{10}$(Unc_Energy_Flux100), $\log_{10}$(Signif_Avg), LP_beta, LP_SigCurv, Variability_Index, and the index of the log parabola spectrum at 500 MeV[5]. The last feature is calculated as the derivative at 500 MeV:

$$n = -\frac{d \log F(E)}{d \log E} \tag{1}$$

where $F(E)$ is the log parabola spectrum of the source reported in the 4FGL-DR3 catalog. These features characterize the main properties of gamma-ray sources, such as the position

[5] The definitions of the feature names can be found in the 4FGL-DR3 catalog [Ab22]

on the sky, the shape of the energy spectrum, overall flux, and variability as a function of time. We have also checked in an earlier analysis that these features have relatively small correlation among themselves, i.e., they are not redundant [BM22].

We split the associated sources into 70% training and 30% testing samples. We show the overall accuracy of classification in the 2- and 3-class cases as a function of maximal tree depth for the RF algorithm or the number of neurons in the NN algorithm in Fig. 1. Each point on this plot is obtained by an average over 10 random splits into training and testing samples. In the RF case, the accuracy saturates around the maximal depth of 10. In the rest of the paper we use the RF algorithm with 50 trees of maximal depth 10. For the NN algorithm, the accuracy saturates around 10 neurons in the hidden layer in the 2-class case or 15 neurons in the 3-class case. In the following analysis we use 15 neurons in the hidden layer.

3 Tests

3.1 ROC curves

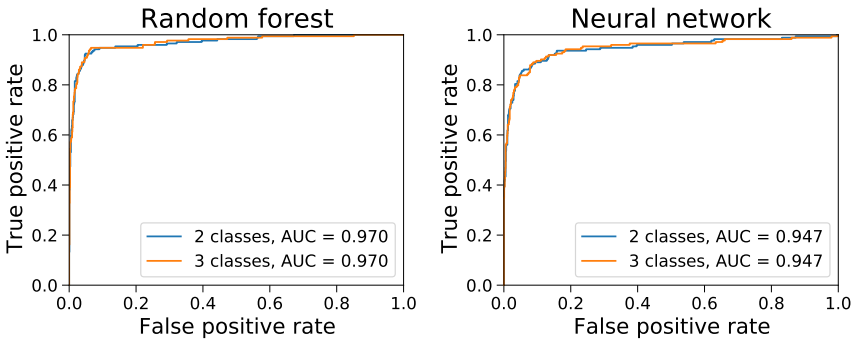


Fig. 2: ROC curves for the Galactic class in the 2- and 3-class classification with RF (left) and NN (right) algorithms. The 3 classes are obtained by dividing the Galactic class into two subclasses: pulsars and non-pulsar Galactic sources. In order to obtain the Galactic class probability in the 3-class case, we sum the probabilities of the two subclasses.

As a first test of the effect of dividing the Galactic class into two sub-classes, we calculated the ROC curve in the 2- and 3-class classifications. Here and in the following sections we use sources from the test sample for the comparison of the 2- and 3-class classifications. In the 3-class case, we sum the probabilities for the two sub-classes of the Galactic class in order to determine the probability for the sources to belong to the Galactic class. We find that there is essentially very little difference in the 2- or 3-class classifications with regards to the 2-class separation (Fig. 2).

3.2 Precision and recall

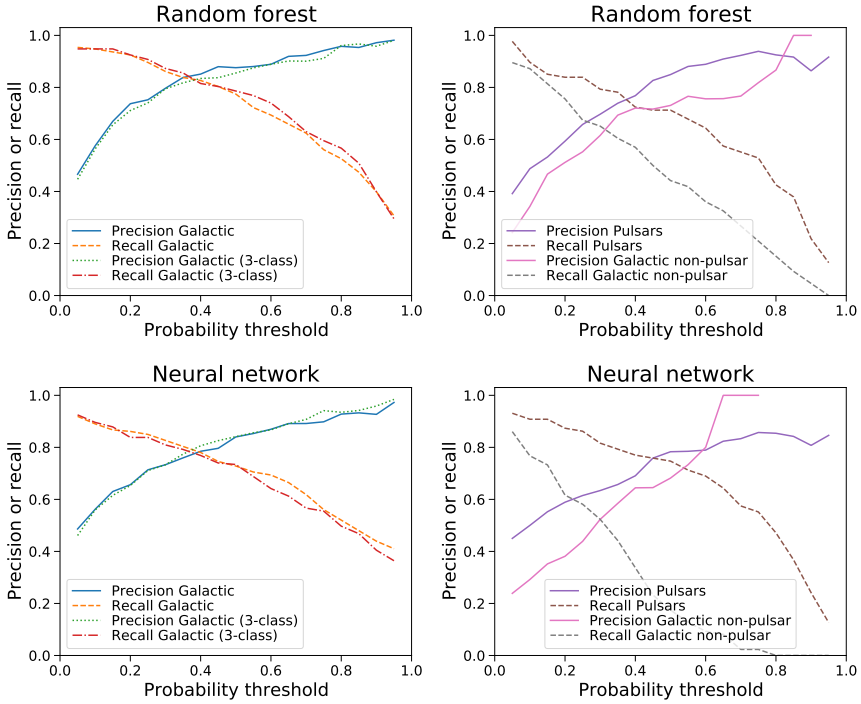


Fig. 3: Precision and recall for 2- and 3-class classifications as a function of probability threshold. Left panel: precision and recall for the 2-class classification (in the 3-class curves, the probabilities for the two sub-classes of the Galactic class are summed to determine the probability for the sources to belong to the Galactic class). The values on the x-axis show the probability above which the sources are classified as Galactic. Right panel: precision and recall for the two sub-classes of the Galactic class (pulsars and non-pulsar Galactic sources). The x-axis on this panel shows the probability above which the sources are classified as respectively pulsars and non-pulsar Galactic sources.

As the next consistency check we calculate the precision and recall for the Galactic class classification. The precision is calculated as the fraction of associated sources, e.g., Galactic, which are also predicted to belong to this class, among all sources predicted to be in this class, while the recall is calculated as the fraction of associated sources predicted to be in the class among all true (associated) sources in this class:

$$\text{precision} = \frac{\#\,\text{true \& predicted}}{\#\,\text{predicted}}; \qquad \text{recall} = \frac{\#\,\text{true \& predicted}}{\#\,\text{true}}. \tag{2}$$

Precision (recall) is connected to the purity (completeness) of a catalog in astronomical observations.

Left panel in Fig. 3 shows the precision and recall for the 2-class classification. We compare the direct training of RF with the 2 classes to training with 3 classes, where we sum the probabilities of the Galactic class sub-classes in order to obtain the Galactic class probabilities. Similarly to the ROC curves, there is very little difference between training with 2 and 3 classes for the 2-class precision and recall. We show precision and recall for the two sub-classes of the Galactic class on the right panel of Figure 3. For example, at the probability threshold of 0.5 the precision for pulsars and Galactic non-pulsar sources is 0.85 and 0.73 respectively. The recall for pulsars and Galactic non-pulsar sources at the probability threshold of 0.5 is respectively 0.71 and 0.44. This can be compared with the precision and recall for the Galactic class in the 2-class classification (training with 2 classes) of 0.88 and 0.77 respectively. We find that the fraction of recovered pulsars in the 3-class classification (71%) is slightly lower than the fraction of all Galactic sources recovered in the 2-class classification (77%), while the fraction of associated pulsars among all sources predicted to be pulsars in the 3-class case (85%) is comparable to the fraction of associated Galactic sources among all sources predicted to be Galactic in the 2-class case (88%). In addition to that, about half (44%) of Galactic non-pulsar sources can be recovered in the 3-class classification, while the fraction of associated Galactic non-pulsar sources among the predicted ones is 73%. Overall, we find that the 3-class classification provides additional information about the sub-classes of the Galactic class without reduction in precision and recall of the parent class.

3.3 Reliability diagrams

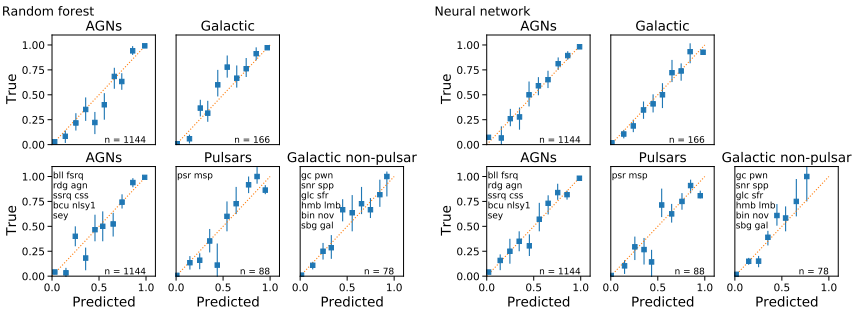


Fig. 4: Reliability diagrams for the 2- and 3-class classification for RF (left) and NN (righ) algorithms. We show the 4FGL-DR3 labels of the classes in the upper left corners in the 3-class classification (bottom panels). The total numbers of sources in the test samples are shown in the lower right corners. The red dotted lines show perfect calibration: y = x.

The last check that we performed is the calculation of reliability (or calibration) diagrams for all classes in the 2- and 3-class classification cases. The top panels in Fig. 4 show the reliability diagrams for AGNs and Galactic classes in the 2-class case, while the bottom panels of the figure shows the reliability diagrams for AGNs, pulsars, and Galactic non-pulsar classes in the 3-class case. The x-axis shows bins in predicted probability: the output score

of the RF (left panels) and NN (right panels) algorithms. The x values of the points are computed as the average predicted probability in the bin. The y values are computed as the number of the associated sources in the corresponding class divided by the total number of sources in the probability bin. The error bars are estimated using 68% containment of the binomial distribution with the total number of sources given by the number of sources in the probability bin and the probability equal to the center of the bin (divided by the total number of sources in the bin in order to obtain values between 0 and 1).

Technically the output score of the ML algorithms does not need to have a probabilistic interpretation and can be corrected using the calibration diagrams in order to be interpreted as a probability for a source to belong to a certain class. Nevertheless, it is reassuring that in both the 2- and 3-class cases the RF and NN scores are consistent with the probabilistic interpretation without additional corrections (within the uncertainty bands estimated from the binomial distribution).

4 Conclusions and Discussion

In the paper we performed a classification of Fermi-LAT gamma-ray sources into two or three classes. Each of these classes encompasses several physical classes reported in the 4FGL-DR3 Fermi-LAT catalog [Ab22]. We use RF and NN methods for the classification of the sources. In the calculations we use the class probabilities estimated by the algorithms (rather than the class predictions). The three classes have an hierarchical structure: two of the three classes are obtained by sub-dividing one of the classes in the 2-class case, while the third class is the same as in the 2-class case. One of the main motivations of the work was to study how such subdivision of classes affects the 2-class classification performance (in the 3-class case the probabilities of the classes obtained by the subdivision of one of the classes in the 2-class case are summed in order to obtain the probability of the parent class).

We find that, at least in the case of 2- and 3-class classification of Fermi-LAT gamma-ray sources with the RF and NN algorithms, the performance measured by the ROC curves, precision, and recall of the 2-class classification is not degraded by subdivision of one of the classes. The 3-class classification provides more information about the possible types of sources than the 2-class case, which is valuable for probabilistic classification of unassociated sources. Additionally, we determine the reliability diagrams and show that both in 2- and 3-class cases the output score of the RF and NN algorithms is consistent with the interpretation as a class probability. Consequently, the probabilities estimated by the NN and RF algorithms reflect the true fraction of the sources belonging to each class, including low probabilities. Thus, one can use the sum of probabilities as an estimate of the number of sources in a class in population studies, which include both associated and unassociated sources.

The results obtained in this work show that further subdivision of classes may be possible without degrading performance of classification of the parent classes. In this paper we

discuss an example of the physically motivated subdivision of classes by dividing Galactic sources into pulsars and other Galactic sources. In general, there are many ways that the 23 physical classes of sources can be separated into 2, 3, 4 or more groups of sources. If the physical classes were well separated in the feature space, then it would not be necessary to combine classes into larger groups, as one could perform the classification with the 23 physical classes. Provided that there is a different degree of overlap among the classes and that some classes have few members, a physically motivated determination of groups of classes may not be optimal from the point of view of classification. In general, one can also use either unsupervised (e.g., Gaussian mixture model) or supervised (e.g., RF or NN) ML algorithms in order to determine separation of physical classes into a few larger groups, which provide an optimal classification. Such determination of optimal groups of physical classes will be the subject of a follow up analysis.

The question of effects of subdivision of classes on classification performance is not limited to the classification of gamma-ray sources, but may have important applications in any classification problem where an hierarchical definition of classes is possible. In the example case of the gamma-ray sources considered in this paper, the subdivision of a class does not degrade the performance of the classification of the parent class, while it provides additional information about the subclasses. Thus, in view of application to other classification problems and for generalizations to multi-class classification, the question of effects of subdivision of classes on classification performance deserves a further, more detailed study.

References

[AAA+12] Ackermann, M.; Ajello, M.; Allafort, A., et al.: A Statistical Approach to Recognizing Source Classes for Unassociated Sources in the First Fermi-LAT Catalog. ApJ 753/1, p. 83, July 2012, arXiv: 1108.1202 [astro-ph.HE].

[AAA+15] Acero, F.; Ackermann, M.; Ajello, M., et al.: Fermi Large Area Telescope Third Source Catalog. ApJS 218/, p. 23, June 2015, arXiv: 1501.02003 [astro-ph.HE].

[AAA+20] Abdollahi, S.; Acero, F.; Ackermann, M., et al.: Fermi Large Area Telescope Fourth Source Catalog. ApJS 247/1, p. 33, Mar. 2020, arXiv: 1902.10045 [astro-ph.HE].

[Ab22] Abdollahi, S. et al.: Incremental Fermi Large Area Telescope Fourth Source Catalog. arXiv:2201.11184/, Jan. 2022.

[BM22] Bhat, A.; Malyshev, D.: Machine learning methods for constructing probabilistic Fermi-LAT catalogs. A&A 660/, A87, Apr. 2022, arXiv: 2102.07642 [astro-ph.HE].

[FKM21] Finke, T.; Krämer, M.; Manconi, S.: Classification of Fermi-LAT sources with deep learning using energy and time spectra. MNRAS 507/3, pp. 4061–4073, Nov. 2021, arXiv: 2012.05251 [astro-ph.HE].

[LP17] Lefaucheur, J.; Pita, S.: Research and characterisation of blazar candidates among the Fermi/LAT 3FGL catalogue using multivariate classifications. A&A 602/, A86, June 2017, arXiv: 1703.01822 [astro-ph.HE].

[Lu20] Luo, S.; Leung, A. P.; Hui, C. Y.; Li, K. L.: An investigation on the factors affecting machine learning classifications in gamma-ray astronomy. MNRAS 492/4, pp. 5377–5390, Mar. 2020, arXiv: 2001.04081 [astro-ph.IM].

[Mi16] Mirabal, N.; Charles, E.; Ferrara, E. C.; Gonthier, P. L.; Harding, A. K.; Sánchez-Conde, M. A.; Thompson, D. J.: 3FGL Demographics Outside the Galactic Plane using Supervised Machine Learning: Pulsar and Dark Matter Subhalo Interpretations. ApJ 825/1, p. 69, July 2016, arXiv: 1605.00711 [astro-ph.HE].

[Pe11] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12/, pp. 2825–2830, 2011.

[Sa16] Saz Parkinson, P. M.; Xu, H.; Yu, P. L. H.; Salvetti, D.; Marelli, M.; Falcone, A. D.: Classification and Ranking of Fermi LAT Gamma-ray Sources from the 3FGL Catalog using Machine Learning Techniques. ApJ 820/1, p. 8, Mar. 2016, arXiv: 1602.00385 [astro-ph.HE].

[ZKZ21] Zhu, K.-R.; Kang, S.-J.; Zheng, Y.-G.: Searching for AGN and pulsar candidates in 4FGL unassociated sources using machine learning. Research in Astronomy and Astrophysics 21/1, p. 015, Jan. 2021, arXiv: 2001.06010 [astro-ph.HE].

Astronomical Image colorization and up-scaling with Conditional Generative Adversarial Networks

Shreyas Kalvankar[1], Hrushikesh Pandit[2], Pranav Parwate[3], Atharva Patil[4], Snehal Kamalapur[5]

Abstract: This research aims to provide an automated approach for the problem of Image colorization and Single Image Super Resolution by focusing on a very specific domain: astronomical images, using Generative Adversarial Networks. We explore the usage of various models in RBG and L*a*b color spaces. We use transfer learning owing to a small data set, using pre-trained ResNet-18 as a backbone encoder and fine-tune it further. The model produces visually appealing images that are high resolution and colorized. We present our results by evaluating the GANs quantitatively using distance metrics such as L1 distance and L2 distance in each of the color spaces across all channels to provide a comparative analysis. We use Fréchet inception distance (FID) to compare the distribution of the generated images and real image to assess the model's performance.

Keywords: Astronomical images; Generative Adversarial Networks; Image Colorization; Single Image Super Resolution

1 Introduction

Automated colorization of grey scales images has been subjected to much research in the machine learning communities. It is an area of research that has untapped potential in applications: from black and white photo reconstruction, image augmentation, image enhancement to video restoration for improved interpretability.

The principal objective of super resolution imaging is to transform a low resolution image into a high resolution one based on a set of low-resolution images to rectify the limitations that existed during the procurement of the original images, ensuring better visualization and recognition for scientific and non-scientific purposes. Even the best up-scaling algorithms are unable to effectively reconstruct non-existing data. Conventional methods rely on low-information and a smooth interpolation between known pixels. These methods can effectively be treated as a convolution with a kernel which encodes no information about the original image. GANs can be used to hallucinate high frequency data in a super scaled image that does not exist in the smaller image. By using the above mentioned method,

[1] K. K. Wagh Institute of Engineering Education and Research, Department of Computer Engineering, India
shreyas.kalvankar@kkwagh.edu.in
[2] hrushikesh.pandit@kkwagh.edu.in
[3] pranav.parwate@kkwagh.edu.in
[4] atharva.patil@kkwagh.edu.in
[5] smkamalapur@kkwagh.edu.in

doi:10.18420/inf2022_40

instead of a perfect reconstruction, a rather plausible guess can be made at what the lost data might be, constrained to reality by a loss function which penalizes deviations from the ground truth image.

As noted in Peh; Marshland [PM19], a huge number of raw images are unprocessed



(a) Ground Truth (b) ResNet18 Full trained L*a*b

Fig. 1: Results of fine-tuned ResNet18 in L*a*b colorspace compared with ground truth images

and unnoticed in the Hubble Legacy Archives. These raw, low resolution and grey scale images are unfit to be shared with the world. While the processing of these images is already time consuming despite human intervention, it is only poised to get worse with newer advancements. This calls for the automation of colorization and super resolution of the images in Hubble dataset, which would make it easier for astronomers to visually identify, analyze and discard images from the Hubble Dataset.



(a) Original image (b) GAN output

Fig. 2: Fig. 2a is the original image and Fig. 2b is the image produced by our GAN

2 Literature Review

2.1 Image Colorization

2.1.1 Deep Colorization

The recent advances in Convolutional Neural Networks have made them a de facto standard for most image problems. CNNs are peculiar in their ability to learn and differentiate colors, patterns and shapes within an image and their ability to associate them with different classes.

Cheng et al. [CYS16] proposed a per pixel training for neural networks using DAISY [TLF08], and semantic [LSD15] features to predict the chrominance value for each pixel, that used bilateral filtering to smooth out accidental image artifacts. With a large enough dataset, this method proved to be effective with a simple Euclidean loss function against the ground truth values. Finally, Dahl [Da16] successfully implemented a system to automatically colorize black & white images using several ImageNet-trained layers from VGG-16 as proposed in Simonyan; Zisserman [SZ15] and integrating them with auto-encoders that contained residual connections. These residual connections merged the outputs produced by the encoding VGG16 layers and the decoding portion of the network in the later stages.

2.1.2 Generative Adversarial Networks

Goodfellow et al. [Go14] introduced the adversarial framework that provides a different training approach using generative distribution of $p_g(x)$ over the input data x. Since its inception in 2015, many extended works of GAN have been proposed including DCGAN [RMC16], Conditional-GAN [MO14], iGAN [Zh18], Pix2Pix [Is18]. Radford et al. [RMC16] applied the adversarial framework for training CNNs as generative models for images, demonstrating the viability of *deep convolutional generative adversarial networks*. Instead of generating images from random noise, Conditional-GAN [MO14] uses a condition to generate output image. GANs can be enlarged into a conditional model by adding some extra information 'y' (i.e.the condition) as an additional layer to both the generator and the discriminator. For e. g. a grayscale image is the condition for colorization of an image. Pix2Pix [Is18] is a Conditional-GAN with images as the conditions. The network can learn a mapping from input image to output image and also learn a separate loss function to train this mapping. Pix2Pix is considered to be the state of the art architecture for image-image translation problems like colorization.

2.2 Image Upscaling

2.2.1 The interpolation based SR image approach

The interpolation-based SR approach constructs a high resolution image by casting all the low resolution images to the reference image and then combining all the information available from every image available. Most known Bayesian-based SR approaches are maximum likelihood (ML) estimation approach and maximum a posterior (MAP) estimation approach. While Tom; Katsaggelos [TK96] proposed the first ML estimation based SR approach with the aim to find the ML estimation of high resolution image, some proposed a MAP estimation approach. MAP SR tries to takes into consideration the prior image model to reflect the expectation of the unknown high resolution image.

2.2.2 Super Resolution - Generative Adversarial Networks (SR-GAN)

Ledig et al. [Le17] introduced SRGAN in 2017, which used a SRResNet to upscale images with an upscaling factor of 4x. SRGAN is currently the state of the art on public benchmark datasets.

3 Methodology

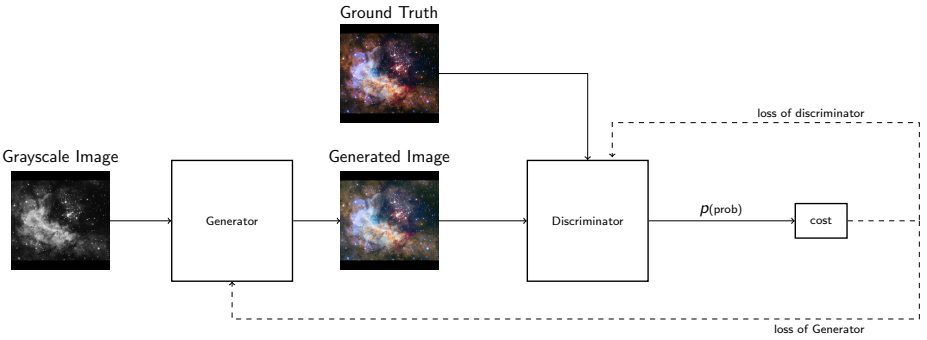


Fig. 3: Approach for image colorization using conditional GANs. A sample grayscale image is input to the generator and the generated image along with the ground truth is fed to the discriminator. p is the probability that the discriminator thinks the two images are similar.

3.1 Data collection

The data is scraped off the Hubble Legacy Archive[6]. The scraper tool which was used, courtesy of Peh; Marshland [PM19][7], scraped off hundreds of thousands of colorized images the archive has available. A filter for M101 (Messier 101) galaxy rendered more than 80 thousand images with a 1 degree difference between consecutive right ascensions. We also scraped the Hubble Heritage project which releases the highest-quality astronomical images that are stitched together, colorized and processed to eliminate noise. There were $\sim$ 150 of these images that were actually useful. We then scraped images from the main Hubble website. This provided another $\sim$ 1000 images approx. Limited by our computational resources, we used images in dimensions of 256×256 pixels with RGB color channels.

[6] Based on observations made with the NASA/ESA Hubble Space Telescope, and obtained from the Hubble Legacy Archive, which is a collaboration between the Space Telescope Science Institute (STScI/NASA), the Space Telescope European Coordinating Facility (ST-ECF/ESA) and the Canadian Astronomy Data Centre (CADC/NRC/CSA).

[7] GitHub repository: https://github.com/KMarshland/hla-scraper

3.2 Transferred learning and model tweaking

We follow the method proposed in Isola et al. [Is18], with a few minor tweaks for significantly reducing the amount of training data and the training time needed to achieve similar results. We use a U-net generator with a pre-trained ResNet-18 as the *encoder*, clipping off half of its layers to obtain the intermediate abstractions. As shown by Ledig et al. [Le17], training the generator separately in a supervised, deterministic manner generalizes the mapping from the input space to outputs. The generator is trained independently over the COCO dataset. Unlike adversarial training, this phase had *mean absolute error* as the cost function. Though trained deterministically, the problem of rectifying incorrect predictions still persists due to constraints over the convergence of the network loss. To combat this, the trained generator was re-trained in an adversarial fashion for further generalization. We hypothesize that re-training in an adversarial fashion will further rectify the subtle color differences that *mean absolute error* couldn't solve.

A pre-trained ResNet-50 with imagenet weights was used as the discriminator with last few layers clipped off. The discriminative network used is something called a "Patch Discriminator"[Is18].This modification to the discriminative network produces a vector of probabilities representing different localities of the input distribution (image) x. Finally, the trained generator and trained discriminator are fine-tuned to fit on our data. These networks are trained in an adversarial fashion using the conditional GAN objective function as proposed by Isola et al. [Is18], with some noise introduced as the L1 norm of generated image tensor to the target image tensor.

3.3 Experimental Setup

We use a mini-batch gradient descent with a batch size of 10, 16 and 32 for different iteration with Adam that has $\beta_1 = 0.5$ and $\beta_2 = 0.999$ as momentum. The generator and discriminator have a learning rate of $2e - 4$ which remains constant throughout the training. We train the model with different epochs ranging between 20,50 and 100, saving the best model weights determined by L1 norm between the output of the generator and the target image. We use early stopping with a patience value of 10. The image size is 256×256.

4 Results and Discussions

4.1 Image Colorization

We present a comparative study of the following models: Basic U-net generator with a residual VGG16 discriminator, hereafter referred to as Basic U-net. A modified U-net with pre-trained ResNet18 as its backbone. We evaluate the model by training it in RGB colorspace to predict 3 *nos.* for every pixel and in L*a*b colorspace where the model predicts

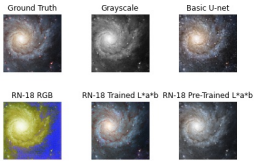


Fig. 4: Results of the colorization study.From top left, (1) the ground truth and (2) input image. The next images belong to the output of following, serially: Basic U-net, ResNet18 in RGB, ResNet18 fine-tuned in L*a*b, ResNet18 pre-trained U-net in L*a*b

the a and b channel alone. We then further fine-tune this model to the task of colorizing astronomical images. Fig. 4 shows that on a particular example, the Basic U-net performs better at predicting results that closely map to the ground truth even if the fine-tuned ResNet-18 shows better results for larger set of inputs. It can also be observed that the Basic U-net architecture does a decent job of faking the sharpness in the original image and that makes the image appear more realistic as compared to rather blurry outputs from the other networks.

We also observe that the model trained in the RGB color space performs poorly at predicting values of all color channels and that results in channel bias (blue in this case). This causes the output, though reconstructed quite accurately, to have a varied color pattern with high emphasis on one color channel. This might be caused by the deep layers which result in diminishing gradients for certain colors over time, causing the model to be strongly biased towards the blue color in this case. An increase in the volume of training data and some random pixel shuffling in forward propagation might solve this problem.

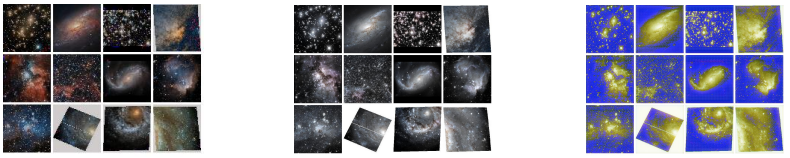
The pre-trained ResNet18 U-net performs decently with the weights gathered by training it over the COCO dataset. The model still lacks the specific coloring intuition in astronomical images and plainly colors specific parts of the images in light colors, leaving majority of the image unaltered. This causes the images to retain grayness. Fig. 6 shows how the other models perform in different color spaces. We observe that the Basic U-net model performs well at predicting the outputs but suffers at predicting the brightness level of pixels. To quantitatively estimate the model performance, we measure the L1 and L2



(a) Sample 1 (b) Sample 2

Fig. 5: Performance of Basic U-net over samples different from the dataset. Left: prediction, right: ground truth

distance between the predictions and the ground truth images in both RGB as well as L*a*b colorspace. Fig. 4 shows that the RGB colorspace isn't ideal for the task of colorization. In terms of the L1 distance, the best performance is achieved by ResNet18 U-net with



(a) Basic U-net (b) ResNet18 pre-trained L*a*b (c) ResNet18 pre-trained RGB

Fig. 6: Results of colorization models (from top to bottom): Basic U-net, ResNet18 U-net pre-trained in L*a*b color space, ResNet18 U-net pre-trained in RGB color space

Model	Color Space	L1 Distance	L2 Distance
ResNet-18 (Pre-trained)	L*a*b	64.5409	2.77
ResNet-18 (Fine-tuned)	L*a*b	65.1119	2.62
ResNet-18 (Pre-trained)	RGB	125.5554	9.04
Vanilla/Basic U-net	RGB	77.3273	3.986

Tab. 1: Colorization: Average per-pixel L1 & L2 distance between generated images and ground truth

pre-trained weights. This goes to prove the unreliability of distance metrics in performance evaluation. The model, though quantitatively, performs better than the finetuned model but in reality, fails to produce visually convincing images. The L2 distance metric shows how the fine-tuned model might, in reality, be fitting slightly better over the data to predict correct color combination as its output. We further investigate the per-channel predictions by the model. Tab. 2 shows the RGB channel averages of the outputs produced by the fine

Distance	Red	Green	Blue
L1 Norm	64.6078	38.5994	92.1283
L2 Norm	3.4531	1.0253	3.8046

Tab. 2: Channel wise averages of ResNet-18 finetuned network

tuned model. It can be observed that the feature embeddings produced by the model in *a, *b colorspaces maps to the green channel with the least error. This could indicate that the model performance is poor on images with high content of red-blue colors. In Tab. 3, L1

Model	L1 Distance		L2 Distance	
	a	b	a	b
ResNet-18 (Pre-trained)	0.00588	0.00501	0.01565	0.03208
ResNet-18 (Fine-tuned)	0.00257	0.00727	0.01721	0.03316

Tab. 3: L1 & L2 channel wise distances in L*a*b colorspace

distance shows the performance of fine-tuned model is better in *a but poor in *b compared to the pre-trained model. The L2 metric rules out the possibility of the fine-tuned model performing better than the pre-trained model. In reality, the fine-tuned model seems better at predicting outputs with the L channel as its input, thus contradicting the quantitative results. It can be concluded from the observations that the fine-tuned model performs better at predicting the Red color channel but suffers in Green and Blue. We use Fréchet inception distance (FID) to evaluate the distribution of generated images with the distribution of real

Model	L1 Distance			L2 Distance			FID
	R	G	B	R	G	B	
ResNet-18 (PT)	69.2401	36.5852	87.7973	3.2890	1.0439	3.9785	66.3568
ResNet-18 (FT)	64.6078	38.5994	92.1283	3.4531	1.0253	3.8046	42.1593

Tab. 4: L1 & L2 channel wise distances in RGB colorspace (PT refers to Pre-trained and FT refers to Fine-tuned)

images. Tab. 4 helps us see that the fine-tuned ResNet-18 U-net achieves the least FID score. This shows that while this model is very adept at hallucinating images it is still not able to predict accurate color values.

4.2 Image Super Resolution

We implement the basic SR-GAN proposed by Ledig et al. [Le17] and train it to improve super-resolution task. We compare the trained model with pre-trained SRGAN model, EDSR-GAN proposed in Lim et al. [Li17] and WDSR-GAN proposed by Yu et al. [Yu18]. Fig. 7 shows that the networks perform excellently, making visual distinction of outputs

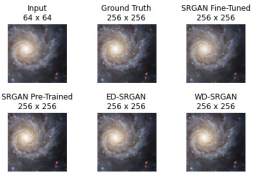


Fig. 7: Results of the super-resolving models. The first two images in the top row are the input image and ground truth respectively. The next images belong to the output of following, serially: Fine-tuned SRGAN, pre-trained SRGAN, EDSR, WDSR

difficult. Upon closer observation, the fine-tuned network seems to produces images slightly better than the counter-parts. The best performing network seems to be the WDSR-GAN. It is evident that the model produces acceptable results on visual inspection. The main



(a) Input Distribution (b) SRGAN fine-tuned output

Fig. 8: Results of SRGAN. Fig. 8a shows the input data and Fig. 8b shows the corresponding output

reason behind this might, again, be the random pixel shuffling between every upscaling pass. As opposed to colorization, super-resolution needs a quantitative estimation to determine which model performs best among the given models. Tab. 5 shows the L1 and L2 norm of predicted results with the ground truth image. It is observed that Ledig's SRGAN, after a bit

Model	L1 Distance	L2 Distance	FID
Ledig SRGAN (Fine-tuned)	87.1090	3.754	75.5873
Ledig SRGAN (Pre-trained)	114.8043	5.953	89.8443
ED-SRGAN	80.9414	3.684	159.5770
WD-SRGAN	79.7262	3.627	127.2920

Tab. 5: Super-Resolution: Per-pixel Average L1 & L2 distance and FID between generated images and ground truth

of fine-tuning delivers good results, in comparison to the the pre-trained version. To further improve this, Lim et al. [Li17] proposed an optimized version of SRGAN by removing the unnecessary modules in the conventional resnets and showed that these Enhanced Deep Residual Networks performed better at upscaling task. When trained in an adversarial manner, the ED-SRGAN performs better than the traditional SRGAN.

Yu et al. [Yu18] further improved the idea by increasing the widening factor ($\times 2$ and $\times 4$) to ($\times 6$ and $\times 9$). This Wide Activation Deep Super Resolution network further improved the performance. When we implement this in an adversarial manner, we achieve excellent results. The FID score listed in Tab. 5 show the lowest FID score is given by the fine-tuned SRGAN, but this just strengthens our previous claim that GANs are particularly good at hallucinating things rather than predicting them.

5 Summary and Conclusion

The colorizing model uses different U-net architectures. A U-net with ResNet-18 as the encoder is implemented and using the pre-trained weights, it is trained on COCO dataset in RGB colorspace. It shows significantly weaker results when compared to the similar network in L*a*b colorspace. The best performing model turns out to be the ResNet18 U-net which is fine-tuned over our dataset to produce results similar to the ground truth.

The Super-resolution model, based on the SRGAN [Le17], performs excellently with pre-trained weights and even better with fine-tuning. We train other state of the art SISR models such as EDSR [Li17] and WDSR [Yu18]. These provide further insights and simultaneously, improve the results. We explore performance metrics and evaluation methodologies of conditional GANs. We find quantitative methods to be unreliable. We base this due to the contradiction of qualitative samples and quantitative measurements of the best performing architecture for colorization models. However, we observe that quantitative estimation is quite reliable for the problem of single image super-resolution to determine which model is better suited for the task.

References

[CYS16] Cheng, Z.; Yang, Q.; Sheng, B.: Deep Colorization, 2016, arXiv: 1605.00075 [cs.CV].

[Da16] Dahl, R.: Automatic colorization, 2016.

[Go14] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y.: Generative Adversarial Networks, 2014, arXiv: 1406.2661 [stat.ML].

[Is18] Isola, P.; Zhu, J.-Y.; Zhou, T.; Efros, A. A.: Image-to-Image Translation with Conditional Adversarial Networks, 2018, arXiv: 1611.07004 [cs.CV].

[Le17] Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; Shi, W.: Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, 2017, arXiv: 1609.04802 [cs.CV].

[Li17] Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K. M.: Enhanced Deep Residual Networks for Single Image Super-Resolution, 2017, arXiv: 1707.02921 [cs.CV].

[LSD15] Long, J.; Shelhamer, E.; Darrell, T.: Fully convolutional networks for semantic segmentation. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Pp. 3431–3440, 2015.

[MO14] Mirza, M.; Osindero, S.: Conditional Generative Adversarial Nets, 2014, arXiv: 1411.1784 [cs.LG].

[PM19] Peh, G. X.; Marshland, K.: Astronomical Image Colorization and Super-resolution using Residual Encoder Networks and GANs, 2019.

[RMC16] Radford, A.; Metz, L.; Chintala, S.: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, 2016, arXiv: 1511.06434 [cs.LG].

[SZ15] Simonyan, K.; Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition, 2015, arXiv: 1409.1556 [cs.CV].

[TK96] Tom; Katsaggelos: Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. In (Anon, ed.): IEEE International Conference on Image Processing. Vol. 2, Proceedings of the 1995 IEEE International Conference on Image Processing. Part 3 (of 3) ; Conference date: 23-10-1995 Through 26-10-1995, IEEE, pp. 539–542, Jan. 1996.

[TLF08] Tola, E.; Lepetit, V.; Fua, P.: A fast local descriptor for dense matching. Proc. CVPR/, June 2008.

[Yu18] Yu, J.; Fan, Y.; Yang, J.; Xu, N.; Wang, Z.; Wang, X.; Huang, T.: Wide Activation for Efficient and Accurate Image Super-Resolution, 2018, arXiv: 1808.08718 [cs.CV].

[Zh18] Zhu, J.-Y.; Krähenbühl, P.; Shechtman, E.; Efros, A. A.: Generative Visual Manipulation on the Natural Image Manifold, 2018, arXiv: 1609.03552 [cs.CV].

Sequential networks for cosmic ray simulations

Pranav Sampathkumar [1], Augusto Antonio Alves Junior [1], Tanguy Pierog [1], Ralf Ulrich for the CORSIKA8 Collaboration

Abstract: A hybrid model of generating cosmic ray showers based on neural networks is presented. We show that the neural network learns the solution to the governing cascade equation in one dimension. We then use the neural network to generate the energy spectra at every height slice. Pitfalls of training to generate a single height slice is discussed, and we present a sequential model which can generate the entire shower from an initial table. Errors associated with the model and the potential to generate the full three dimensional distribution of the shower is discussed.

Keywords: Sequential Neural Networks; Astroparticle Physics; Monte Carlo Simulations

1 Introduction

Cosmic rays are important messengers from the high energy universe. They are also one of the most exciting platforms to study High Energy Physics. They provide us with a window into high energy collisions, which is not possible by current accelerators (of the order $10^{17} eV$). While they provide with extremely high energies, analysing physics using cosmic rays is challenging because of their low intensity and can only be studied using Extensive Air Shower (EAS) which they trigger in the atmosphere. The extensive air showers are a series of cascaded interactions when cosmic rays interact with the earth's atmosphere. Each high energy interaction, triggers a chain reaction, and they produce secondary particle showers. Analysis of such showers is a challenging task because of all the interactions possible in the earth's atmosphere such as, a variety of e/m interactions such as bremsstrahlung, pair production, Compton, Moller, Bhabha processes as well as annihilation and ionization losses. We would also have to simulate similar interactions for high energy muons. All these interactions as well as the demand to be as optimal as possible owing to the large number of times such simulations need to be repeated, for example, in MCMC analysis calls for a demand in more novel approaches to minimize memory overhead and optimize simulations.

The two main approaches currently used are, *thinning* and *hybrid simulations*. In thinning, only a small portion of the shower is explicitly considered for Monte-Carlo simulations while the rest are ignored. The considered showers are re-weighted accordingly. Such an approach, gives rise to artificial fluctuations because of the lower number of particles considered. One way to alleviate this problem is to have a maximum allowable weight to a particle, but this creates limitations on using less detailed sampling and doesn't help speedup the calculation

[1] Institute for Astroparticle Physics, Karlsruhe Institute for Technology, Karlsruhe, Germany
pranav.sampathkumar@kit.edu

doi:10.18420/inf2022_41

process[Ko01]. The second hybrid simulations approach, models the shower using a cascade equation and then after initially running the shower using explicit Monte-Carlo simulations for very high energies. The data is then fed to cascade equations at a threshold. Combining cascade equations and MCMC this way, allows us to get accurate results for the average behaviour of the shower.

The idea of using cascade equations for cosmic ray showers is not new. There have been various attempts at trying to model the longitudinal profile of cosmic ray showers using cascade equations for more than 30 years. The initial shower is usually done explicitly since it involves small number of particles, and it also provides a way to incorporate fluctuations into the simulations since cascade equations are deterministic. We can determine the mean behaviour and the moments of the fluctuations using cascade equations accurately. One of the biggest problems with this approach is, it is limited by our modelling of the cascade equations. Currently, we only have the analytical form of the cascade equations in one dimension. Thus, cascade equations only provide us with the longitudinal profile of the shower. The numerical solution of the cascade, equation can also be avoided by a pre-tabulation of of the secondary showers, using an iterative MC procedure [Al02]. Such pre-tabulation makes the simulations less flexible to changes in the environment and initial conditions.

The current work builds upon the work of CONEX [Be07], in trying to build a neural network model which mimics the behaviour of the analytic form of the cascade equation. Once the network is trained, the network would no longer require any sort of pre-tabulation and is designed to be flexible across various initial conditions. This paper is organised as follows, Sect. 2 serves as an introduction to cascade equations for a general particle shower. Sect. 3 talks about the neural network model. Sect. 4 provides some preliminary results to this approach, and Sect. 5 talks about conclusions and future potential directions for this work.

2 Cascade equations for hybrid simulations

A simple Monte-Carlo process can be described by a linear cascade equation. The current approaches to cascade equations for hybrid simulations are done in a theory driven manner, where our theoretical understanding and ability to derive the analytical form of the cascade equation. Cascade equations in one-dimensional form are derived based on various physics processes, such as hadronic and electromagnetic interactions and they are numerically solved using the tables generated during the explicit simulation as initial conditions.

Various attempts have been made towards cascade equations, with physics parameters,

modelled specifically for hadronic interactions and electromagnetic interactions. The general structure of the cascade equations are as follows.

$$\frac{\partial n_i(E,X)}{\partial X} = \overbrace{\sum_a \int_E^{E_0} \sigma_a n_a(E',X) P_{a\to i}(E,E') dE' - \int_{E_{min}}^{E} \sigma_i n_i(E',X) P_{i\to a}(E,E') dE'}^{\text{Production and Decay Terms}} \underbrace{+S_i(E,X) - \sigma_i n_i(E,X) - \alpha \frac{\partial n_i(E,X)}{\partial E}}_{\text{Source, Interaction and Loss terms}} \tag{1}$$

where, $n_i(E,X)$ is the energy spectra at given depth X. Solution to the cascade equation provides us with the energy spectra $\forall X$. This cascade equation is one dimensional, and helps us calculate the longitudinal profile of the shower.

Production terms in Equation (1), account for the formation of particles by particles at higher energy decaying, and increasing the number of particles at a particular energy bin. *Decay* terms account for the loss of particles at a particular bin by decaying into particles of lower energy. *Source* terms account for sudden appearance and disappearance of particles, not captured by the rest of the cascade equation. It can account for new particles joining the shower during the process of solving it, or particles leaving the cascade equation to be handled differently. *Interaction* and *Loss* terms account for the loss in energy of particles by not performing a explicit cascade beyond a particular energy threshold. This involves ionization losses in the atmosphere, etc.

CONEX solves this system of coupled differential equations, and gets the projection of the shower along the longitudinal axis. One of the problems with extending this approach to higher dimensions is the lack of the analytical form of the cascade equation. In 1D the system is completely described by the energy spectrum at a particular height, but at 3D, we need additional information such as the direction of the subshowers etc in order to fully describe the system. We postulate the existence of such a differential equation which can fully describe the system in 3D though it is not tractable in analytic form. We use a neural network approach where the learning process learns the differential equation which describes the system and can be used to generate the state of the shower at every height.

3 CONEX inspired sequential neural approach

Let us assume that f is the *functional* which takes the energy spectrum at a particular height and gives the energy spectrum at the next height step.

$$n(E, X + \Delta X) = f(n(E,X)) \tag{2}$$

solving the differential equation is equivalent to finding this functional, which helps us iterate through the height and generate the entire shower from an initial condition. The ideology behind the neural network model is that, such a functional can be approximated by the neural network during the training process. This allows us to model the interaction using explicit simulations directly without relying on our understanding of the underlying physics. This also allows us to model higher dimensional distributions since we don't rely on the explicit form of the cascade equations.

As a initial check of the idea, we take a purely electromagnetic cascade in vacuum, and see if it learns the information to perform a single step. The input dataset is generated using CONEX, and we check if given a table at a particular height slice, the network can learn to generate the table at the next height slice. We use a fully connected neural network for this implemented using `PyTorch` [Pa19]. The network has 5 hidden layers, with 512 , 256 , 128 , 256 , 512 nodes each, using MSE as loss function and optimized using ADAM.

4 Preliminary Results

These are the preliminary results for the ability of the network to learn the functional. Fig. 1. shows that the neural network is capable of performing a single step with good accuracy. We find that the neural network learns with around 5% error. When we try to use the learned f, to generate the entire shower, we find that the error soon accumulates and the shower reaches unphysical situations which the network hasn't seen in its training dataset.

A recurrent neural network is ideally suited for such a problem where there are corrective mechanisms for such "time series"data as a hidden layer is passed to the next step. Before using recurrent neural networks, we attempt using sequential neural networks. The neural network is applied iteratively for multiple steps and then the result is back propagated. Such *sequential chains* are useful in helping the network learn the correct physics, so that it doesnt lead to unphysical situations when applied iteratively. The length of the sequential chains is a hyperparameter which needs to be tweaked. The results in this work are done using a sequential chain of ten steps.

Fig. 2 shows the results of using the sequential network to generate the entire shower. We see while the network has higher error earlier in the shower. It has learnt the right physics and has low energy at the later parts of the shower. This is particularly useful given, the later parts of the shower are the ones which are computationally intensive and the part which we hope to replace with our model.

5 Conclusions and Further Work

We see that the neural network can learn the stepping function f, for the one dimensional case. Further, work is needed on analysing the learnt physics by injecting new particles into

the shower in between, a situation the network didn't see in the training dataset. This checks if the network learns that showers can be super imposed. One way to solve this is to hard code the linearity of Eq. (1) into the network to improve the network performance in early parts of the shower. This work is ongoing currently.

After this, the aim is to make the network generate the three dimensional distribution. We also hope that the network can predict the higher moments of the fluctuations apart from the mean behaviour, when the higher moments are provided along with the input distribution. The methods outlined in this work, can also be extended to any simple monte-carlo system since they would be described by a linear cascade equation. We hope this will be a valuable contribution to the cosmic-ray physics community where the showers can be quickly generated for high energies for event-by-event analysis in air-shower arrays or fluorescence light detectors.

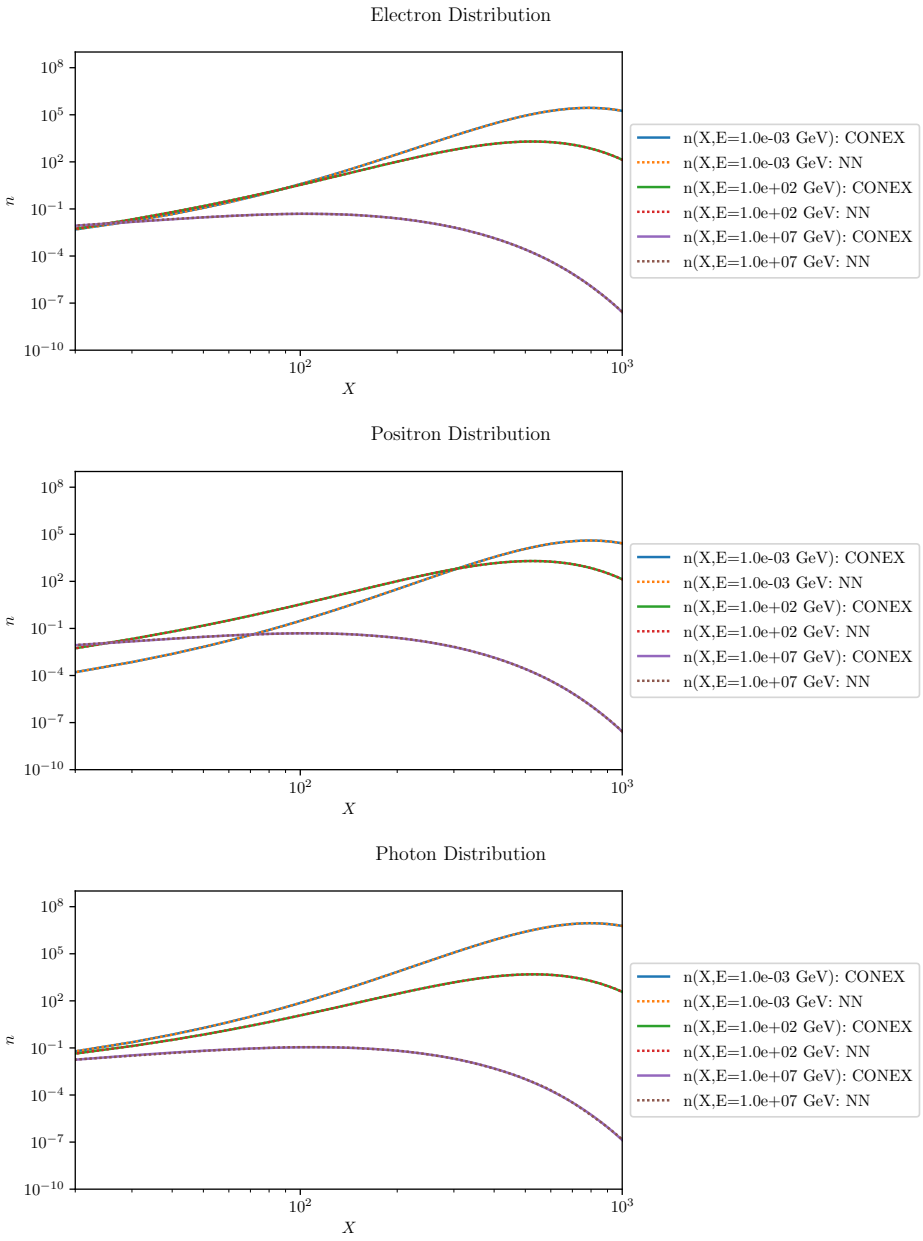


Fig. 1: Results where the neural network performs a single step, the generated showers match with CONEX showers and are accurate to around 5% error.

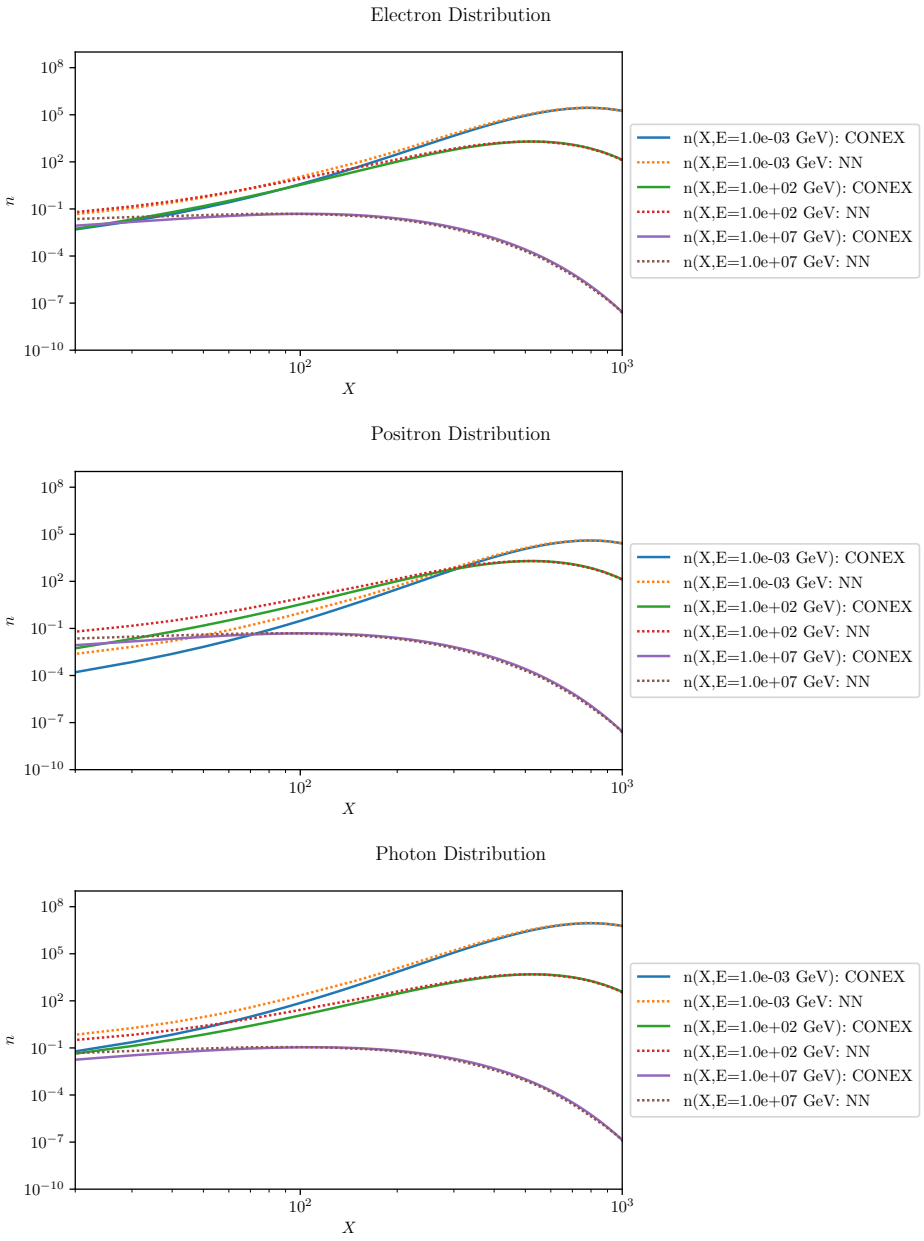


Fig. 2: Results where the neural network generates the entire shower. The generated shower deviates from the CONEX showers earlier in the shower

References

[Al02] Alvarez-Muñiz, J.; Engel, R.; Gaisser, T. K.; Ortiz, J. A.; Stanev, T.: Hybrid simulations of extensive air showers. Phys. Rev. D 66/, p. 033011, Aug. 2002, URL: https://link.aps.org/doi/10.1103/PhysRevD.66.033011.

[Be07] Bergmann, T.; Engel, R.; Heck, D.; Kalmykov, N.; Ostapchenko, S.; Pierog, T.; Thouw, T.; Werner, K.: One-dimensional hybrid approach to extensive air shower simulation. Astroparticle Physics 26/6, pp. 420–432, 2007, ISSN: 0927-6505, URL: https://www.sciencedirect.com/science/article/pii/S0927650506001228.

[Ko01] Kobal, M.: A thinning method using weight limitation for air-shower simulations. Astroparticle Physics 15/3, pp. 259–273, 2001, ISSN: 0927-6505, URL: https://www.sciencedirect.com/science/article/pii/S0927650500001584.

[Pa19] Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Kopf, A.; Yang, E.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library. In (Wallach, H.; Larochelle, H.; Beygelzimer, A.; d'Alché-Buc, F.; Fox, E.; Garnett, R., eds.): Advances in Neural Information Processing Systems 32. Curran Associates, Inc., pp. 8024–8035, 2019, URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

Calculation of the Photon Flux in a Photo-Multiplier Tube with Deep Learning

Jigar Bhanderi[1] Stefan Funk[2] Dmitry Malyshev[3] Naomi Vogel [4] Andreas Zmija[5]

Abstract: Intensity interferometry is part of optical interferometry, which provides a sub-milliarcsecond resolution of astronomical objects. In intensity interferometry one correlates intensities of optical fluxes rather than amplitudes of waves. For a successful measurement one needs a large light collecting area for several telescopes separated by hundreds of meters and good time resolution of the intensity flux. Air Cherenkov telescopes, e.g., H.E.S.S. are a natural candidate for performing such a measurement. One of the important tasks is to determine the rate of photons hitting the PMTs to calculate expectations on the signal-to-noise ratio. For low rates, the individual pulses can be resolved and counted, but for high rates, relevant for the IACTs, the pulses from the photons overlap. We use different neural network algorithms in order to determine the rate of photons hitting the PMT, including the high rates.

Keywords: Intensity interferometry, IACTs, deep neural networks

1 Introduction

Imaging of celestial objects has played a vital role in several remarkable astronomical discoveries and experimental evidences of theories, from upper limit calculations of the angular size of Vega (5 arc-second) by Galileo to creating first images of two supermassive black holes (SMBHs): in the center of M87 galaxy and in the center of the Milky Way galaxy. The imaging of the SMBHs was performed with the radio interferometric measurements by the Event Horizon Telescope (EHT) Collaboration [Co19]. Measurements of the angular diameter of stars is important for testing the theories of stellar structure, evolution, and modeling of the star's atmosphere. Amplitude interferometry with optical telescopes is performed in the ESO Very Large Telescope Interferometer. In optical amplitude intererometry the amplitudes of waves are correlated in real time. This measurement is sensitive to the fluctuations in the atmosphere. As a result, baselines are limited by a few hundred meters. Intensity interferometry (also referred to as the second order interferometry, whereas amplitude interferometry is the first order) is less sensitive to atmospheric fluctuations and provides a possibility of measurements with larger baselines (better angular resolutions) than the amplitude interferometry.

[1] Erlangen Centre for Astroparticle Physics, Erwin-Rommel-Str. 1, Erlangen, Germany jigar.bhanderi@gmail.com
[2] Erlangen Centre for Astroparticle Physics, Erwin-Rommel-Str. 1, Erlangen, Germany s.funk@fau.de
[3] Erlangen Centre for Astroparticle Physics, Erwin-Rommel-Str. 1, Erlangen, Germany dmitry.malyshev@fau.de
[4] Erlangen Centre for Astroparticle Physics, Erwin-Rommel-Str. 1, Erlangen, Germany naomi.vogel@fau.de
[5] Erlangen Centre for Astroparticle Physics, Erwin-Rommel-Str. 1, Erlangen, Germany andi.zmija@fau.de

doi:10.18420/inf2022_42

Intensity interferometry measurements with Imaging Atmospheric Cherenkov Telescopes (IACTs) are performed by measuring the flux from the object of interest, such as a star, with two (or more) telescopes. The fluxes of photons collected by the telescopes are recorded and correlated offline. The amount of correlation depends on the size of the object, flux, and the distance between the telescopes projected on the direction to the object. The flux in each telescope is detected by one or more photo-multiplier tubes (PMTs). Isn't it rather this: In order to estimate the expected signal-to-noise of the correlation, one needs to know the photon rates in the detectors Typical expected photon rates are on the order of a hundred MHz or more, where pulses from individual photons overlap in the measured photo current. As a result, the measurement of the rate by simple counting of pulses is not possible.

In this work, we use deep learning (DL) methods, for a cross-check of the more traditional method of calculating the flux of photons in a PMT based on integrating the charge readout form the PMT. The input data is the readout from the PMT sampled at 1.6 ns timesteps. The output is the rate of photons detected by the PMT.

1.1 Amplitude Interferometry

In astronomy, interferometers are widely used to measure correlation of radiation from the source travelling through different paths. The angular resolution is proportional to λ/b, where λ is the wavelength of radiation and b is separation of the detector elements in the interferometer, i.e., the baseline. Increasing baselines help to improve the resolution for the selected wavelength. There are two types of interferometers: amplitude interferometers (or Michelson's interferometer) and intensity interferometers. In amplitude interferometry, the amplitudes of the electromagnetic waves are added in real time. The detected intensity is the square of the sum of the amplitudes. Provided that the amplitudes of waves are added in real time, a very high optical precision is required so that the difference between light paths is smaller than wavelength of light. In particular, one needs to have a very precise control over the optical paths between the telescopes. Also the atmospheric turbulence introduces random and rapid changes into the paths of light, which can significantly affect interference signal. As a result, the baseline lengths are limited to a few hundreds of meters. Especially to overcome limitation of atmospheric turbulence for stellar surface imaging, several scientific ideas were proposed for space based [Ca07; La09] and ground based telescopes [Cr18]. However, considering different aspects, such as cost, complexity in construction and available alternatives of amplitude interferometry, ground based interferometry has been prioritized. As a result, presently there are several telescopes and collaborations working on optical and infrared interferometric imaging such as Center for High Angular Resolution Astronomy (CHARA) [An20; Br05], GRAVITY [GR17].

1.2 Intensity Interferometry

Intensity interferometry was introduced by Hanbury Brown and Twiss [HT56], who have measured sizes of several stars using Narrabri Stellar Intensity Interferometer [HDA67; HDA74]. The idea is to measure temporal correlations between arrival times or fluxes of photons from two (or more) telescopes, i.e., intensity rather than phase and amplitude. Within certain range of baselines, a fraction of photons have correlated arrival times, while moving the telescopes further results in decorrelation of the arrival times. In intensity interferometry, the second-order correlation function is used as an observable. It can be factorized in spatial and temporal parts:

$$g^{(2)}(\vec{r_1}, r_2, \vec{\Delta t}) = \frac{\langle I(\vec{r_1}, t) I(\vec{r_2}, t + \Delta t) \rangle}{\langle I(\vec{r_1}, t) \rangle \langle I(\vec{r_2}, t) \rangle} = 1 + g^{(2)}(\vec{r_1} - \vec{r_2}) g^{(2)}(\Delta t) \tag{1}$$

where the spatial correlation is proportional to the Fourier transform of emission profile of the source, while for the same spatial position ($\vec{r_1} = \vec{r_2}$), the interference is reduced to temporal correlation, which is related to the Fourier transform of the observed optical frequency spectrum $S(\nu)$. The signal to noise ratio is [Zm20]

$$S/N = C \epsilon n(\nu) \sqrt{\frac{T}{\tau_e}} \tag{2}$$

where, S is the integrated correlation signal, N is statistical fluctuations, C is the telescope collection, ϵ is the system's detection efficiency, $\eta(\nu)$ is the the photon flux, T is the observation time, and τ_e is system's time resolution. Thanks to recent progress in electronics and computing, which allows sampling of the current from PMTs with steps of $\sim$ 1 ns, storage and processing of the large amounts of data, a new era in intensity interferometry with Cherenkov telescopes has begun [22; Ac20]. In this work, we focus on one of the aspects of the intensity interferometry measurements with IACTs: the calculation of the flux necessary for the estimates of the signal to noise ratio in Equation (2). Provided that the expected photon rate is very high, so that direct counting of peaks in the PMT readout is not possible, we develop DL methods for estimate the rate of photons detected by the PMTs.

1.3 Deep Learning

Machine learning methods can provide predictive models to solve complex problems either with simulated or measured data, using classification or regression approach. In the last decade, DL has been extensively used to solve various problems such as Natural Language processing [Su14], speech recognition [Hi12] and Computer Vision [He16; LBH15; Su12]. Especially, Computer Vision methods have been used thoroughly in physics [Ch20; PP18]. In this work, we focused on various types of DL models such as Convolutional Neural Network (CNN), Long-Short Term Memory Unit (LSTM) and Gated Recurrent Unit (GRU) in order to determine the photon rates detected by PMTs. We compared the results with

more traditional methods of estimating the photon rate: peak-finding method from scipy[6] library and charge integration [21; Zm21]. The peak finding method provides good results at low rates and underestimates the number of photons at high rates, when many peaks are overlapping.

2 Data Preparation and model architecture

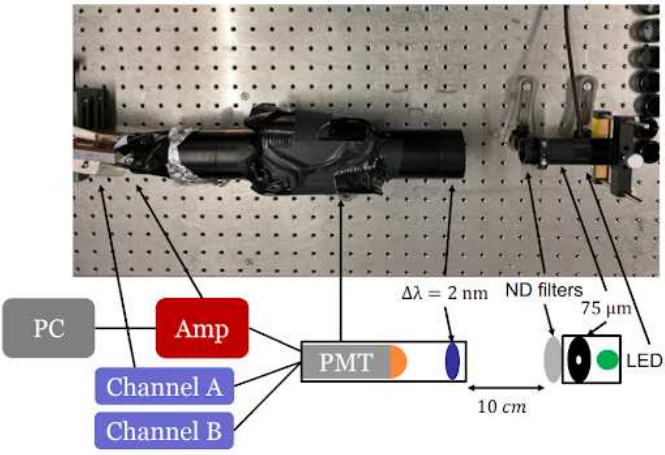


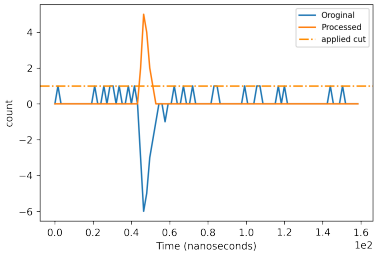
Fig. 1: Schematic diagram of experimental setup [20].

Fig. 1 The setup of the laboratory interferometer which we used for photo current measurements. During the experiment we used combination of neutral density filters (grey filters) [20] to generate samples with different photon rates. The readout from the PMT is a time-series with 1.6 ns steps. In the analysis we applied the following steps: extract photon pulse shapes from time-series from PMT, then create Monte-Carlo simulation data to train and test model. Then use the model trained on MC data for prediction over unseen data from the laboratory experiment. Use the procedure developed with the lab measurements for the measurement of the flux from Sirius, which were taken using 30nm band-width filter with IceACT telescope of 0.5 m diameter.

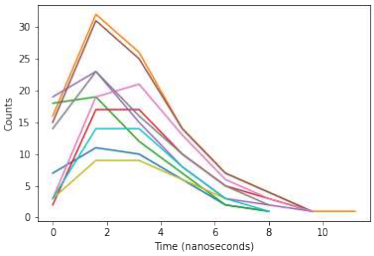
Obtained time series contains pulse shapes, and it is negatively directed. First we flipped up entire time series with respect to baseline. Then, from baseline, we considered smallest positive count height as baseline noise and applied cut at this selected threshold (Fig. 2a). The original and preprocessed data are shown by blue and orange lines in Fig. 2a respectively.

The next important step is pulse shape extraction. We applied preprocessing (mentioned above) on time-series with lowest transmittance, because in such data, individual pulse shapes are likely to be complete and without any overlapping with other pulse shapes. We

[6] https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.find_peaks.html



(a) Preprocessing of the data. Blue line – raw signal from the PMT. Orange line – preprocessed data.



(b) Examples of signal profiles induced by a single photon in the PMT.

Fig. 2: Data preprocessing and PMT response to a single photon.

breakdown processed time series into vectors of 100 time-steps, from which we removed samples, which contain a partial pulse shape at an edge of the vector and those without any pulse shapes. The remaining set have 0.4 Million complete pulse shapes (Fig. 2b), which we splitted into training,validation, and testing sets with 6:2:2 ration. Each set of shapes was used for the creation of MC data at the next step.

Used MC-algorithm, requires four type of inputs:1.) desired sample size 2.) required range of photon numbers should be added in a sample 3.) number of examples for all selected range 4.) set of extracted pulse shapes. We used two different selection criteria. First, sample size 100, rates from 0 to 60. Selection of examples for MC training, testing and validation set is 10000, 10000 and 5000. In another set, we increased sample size selection to 1000 and range selection (0 to 160), while decreased number of examples for each rate, so that file size would not be extremely large. In both cases, labels are single integer quantity related to each sample.

2.1 Model architectures

We used CNNs LSTM (bidirectional) and GRU (bidireactional) without regularization. At the output we have a single neuron with ReLu activation that predicts the photon rate. We used 3 types of CNN architectures. First, two consists only two Convolutional layers of order of thickness 4-2 (CNN-1,Model-1) and 4-32 (CNN-2,Model-2), while other is CNN variant ResNet26 (CNN-3,Model-3)[7]. Our selection of unit cell in one direction is 128 for LSTM and GRU. We used Mean Squared Error as loss function and Adam optimizer. For CNN models we used smaller learning rate (10^{-4}) and 128 mini-batch size, while for LSTM and GRU we used learning rate 10^{-3} and 256 mini-batch size. We used Tensorflow framework

[7] The architecture was inspired by the exercise in a course from Coursera `https://www.coursera.org/lecture/custom-models-layers-loss-functions-with-tensorflow/understanding-residual-networks-Dv4Ub`. Here we used MaxPooling layer instead of AveragePooling, and removed BatchNormalization layers.

for training and testing, by selecting ReLu activation functions for CNN models, while for sequential models we used default activation functions. Every time we trained our model for 50 iterations. Our MC algorithm and ResNet model are mentioned on github repository[8].

3 Results

To compare performances of different methods, we first used MC data with sample size 100 for training models. Then we performed two type of testing. First, testing of models over MC test dataset where ground truth is known, and second, over unseen measurements from laboratory where labels are unknown but can be estimated form transparencies of grey filters. Fig. 3 shows the comparison of predictions from different methods over MC test dataset.

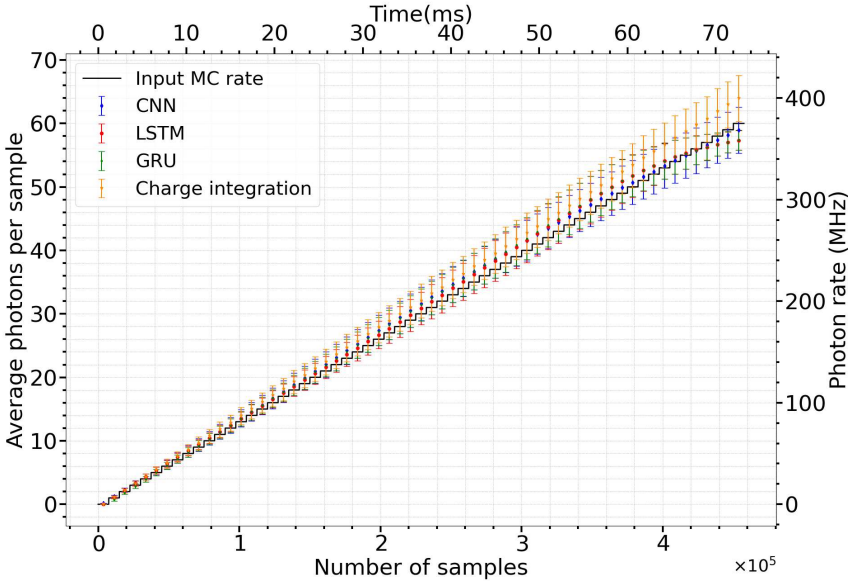


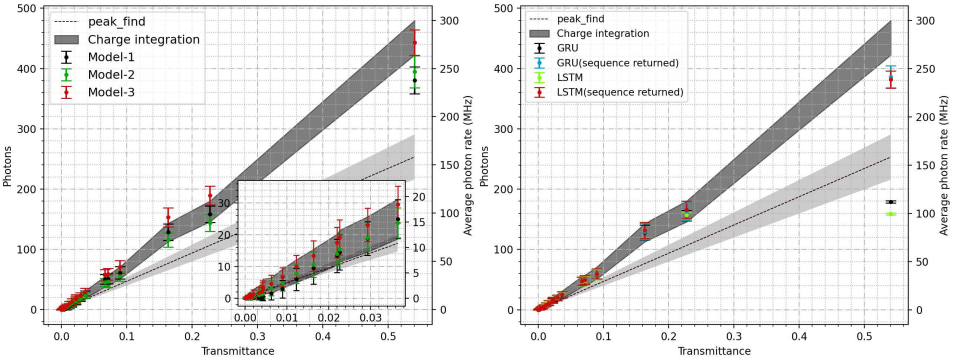
Fig. 3: The number of photons in the input data vectors of 100 steps is shown on the y-axis (e.g., 50 means that there are 50 photons detected during the 100 time steps, where each step is 1.6 ns). x-axis shows the number of samples simulated with a particular number of photons ordered according to the number of photons. For the CNN model (blue points) we use ResNet architecture.

Up to 350MHz all methods predictions are consistent with ground truth. Specially DL models are predicting close to input MC rate even at higher rate.

Further, using MC data with larger sample size 1000, we trained our model with the same configurations and tested over grey filter measurements with higher transmittance.

[8] https://github.com/jigar2099/Monte-Carlo/tree/main

According to expectations, statistical uncertainty of each model has reduced. Dark grey band in Fig. 4 is prediction band of the Charge Integration method, while light grey band is extrapolation of first six data points with 1σ uncertainty from the peak finding method. Fig. 4a represents comparison of predictions from CNN models, in which model predictions are consistent with CI. Further we have tried two different configurations of LSTM and GRU models. In the first one, sequences are returned .i.e, we extract output from each unit cell. In the second one the output is extracted from final unit cell. Fig. 4b represents prediction comparison between two configurations of LSTM, GRU models with CI and the extrapolation of the peak finding estimation from low rates. The DL estimates are consistent with CI method for all frequencies considered in this test, while the extrapolation of the peak finding calculation is systematically lower than both DL and CI for rates above 50 MHz.



(a) Prediction comparison of CNN models with sample size 1000

(b) Prediction comparison of LSTM and GRU with sample size 1000

Fig. 4: Model-1 and Model-2 are similar to CNN-1 and CNN-2, while Model-3 is ResNet26. In GRU (bidirectional) and LSTM (bidirectional), last cell's output is given to final dense neuron, while in sequence returned configurations output from each unit is extracted and used at final layer.

After that, we created MC data from Sirius measurements with sample size 1000, and trained our models with the same configurations as we used for grey filter measurements. In predictions we used samples of size 1000. Fig. 5 represents prediction comparision of our models for Sirius measurements. Predictions from our models are consistent with each other. Changes in the rate in Fig. 5 during the 50 mins of observations are due to the movement of Sirius in the sky. One sharp bump at the end is due to Sirius moving behind a tree near the horizon.

4 Conclusion

From our experiments, DL models, trained over large data set, have predicted photon rates consistent with the more traditional method of charge integration. Moreover, DL models trained over a larger sample size tend to provide predictions with smaller uncertainties as expected from testing them over MC test dataset. We find that models with fewer parameters

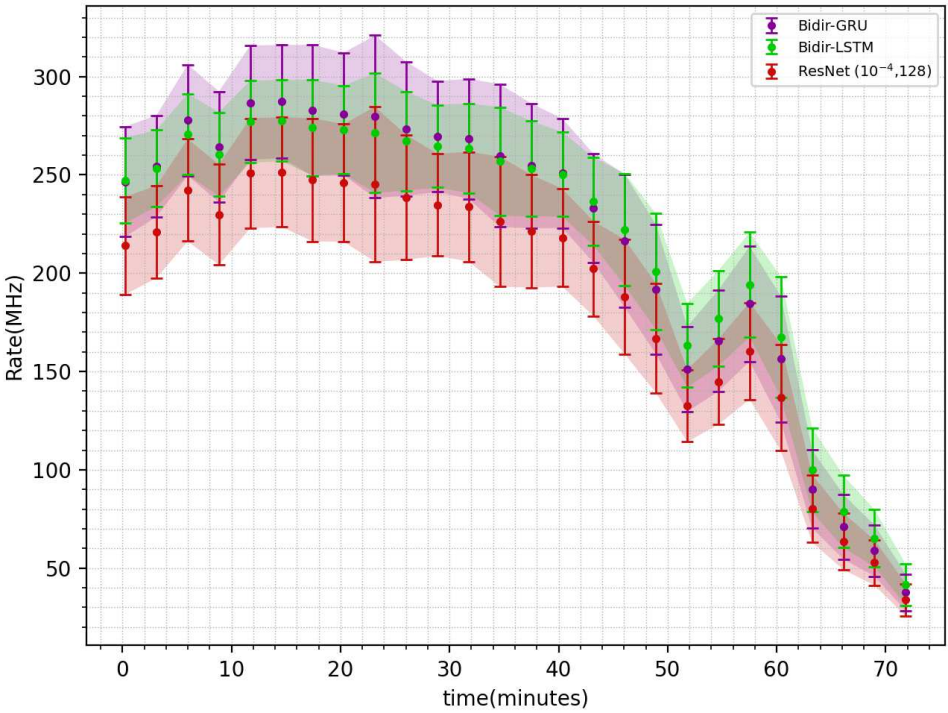


Fig. 5: Prediciton comparison between Bidirectional LSTM and GRU models with returned sequence configuration and ResNet26

(CNNs) as well as more complex models, such as Bidireactional LSTM and GRU, are able to predict with similar accuracy as the charge integration method. Implementation of our models for Sirius measurements is in agreement with lab measurements. Additionally, the predictions of the models are consistent within the uncertainty bands of each other. At this point, DL approach has potential to provide a valuable cross-check for the more traditional methods of the estimation of photon rate measured by a PMT. The ground truth for lab measurement (e.g., derived with a set of grey filters) is required for the confirmation of the flux calculation for astronomical sources.

References

[20] Calibration of Photomultiplier Tubes for Intensity Interferometry at H.E.S.S./, 2020, URL: https://ecap.nat.fau.de/wp-content/uploads/2021/01/2020-MA-NVogel.pdf.

[21] Photon flux calculation using Deep Learning./, 2021, URL: https://ecap.nat.fau.de/wp-content/uploads/2022/03/2021_Bhanderi_MA_thesis.pdf.

[22] VERITAS - Very Energetic Radiation Imaging Telescope Array System./, 2022, URL: https://veritas.sao.arizona.edu/.

[Ac20] Acciari, V. A.; Bernardos, M. I.; Colombo, E.; Contreras, J. L.; Cortina, J.; Delgado, C.; Díaz, C.; Fink, D.; Hassan, T.; Jiménez-Martínez, I.; Lyard, E.; Mariotti, M.; Mangano, S.; Mirzoyan, R.; Produit, N.; Polo, M.; Schweizer, T.; Walter, R.; Will, M.; Wunderlich, C. W.; Schweizer, T.; Zampieri, L.: Intensity interferometry with MAGIC. In (Tuthill, P. G.; Mérand, A.; Sallum, S., eds.): Optical and Infrared Interferometry and Imaging VII. Vol. 11446, International Society for Optics and Photonics, SPIE, pp. 346–359, 2020, URL: https://doi.org/10.1117/12.2561559.

[An20] Anugu, N. et al.: MIRC-X: A Highly Sensitive Six-telescope Interferometric Imager at the CHARA Array. The Astronomical Journal 160/4, p. 158, Sept. 2020.

[Br05] ten Brummelaar, T. A. et al.: First Results from the CHARA Array. II. A Description of the Instrument. The Astrophysical Journal 628/1, pp. 453–465, July 2005.

[Ca07] Carpenter, K. G.; Lyon, R. G.; Schrijver, C.; Karovska, M.; Mozurkewich, D.: Direct UV/optical imaging of stellar surfaces: the Stellar Imager Vision Mission. In. Vol. 6687, International Society for Optics and Photonics, SPIE, pp. 171–182, 2007, URL: https://doi.org/10.1117/12.732440.

[Ch20] Chianese, M. et al.: Differentiable strong lensing: uniting gravity and neural nets through differentiable probabilistic programming. Monthly Notices of the Royal Astronomical Society 496/1, pp. 381–393, May 2020.

[Co19] Collaboration, T. E. H. T. et al.: First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole. The Astrophysical Journal Letters 875/1, Sept. 2019, URL: https://iopscience.iop.org/article/10.3847/2041-8213/ab0ec7.

[Cr18] Crouzet, N. et al.: Four winters of photometry with ASTEP South at Dome C, Antarctica. A&A 619/, A116, 2018, URL: https://doi.org/10.1051/0004-6361/201732565.

[GR17] GRAVITY Collaboration et al.: First light for GRAVITY: Phase referencing optical interferometry for the Very Large Telescope Interferometer. A&A 602/, A94, 2017, URL: https://doi.org/10.1051/0004-6361/201730838.

[HDA67] Hanbury Brown, R.; Davis, J.; Allen, L. R.: The stellar interferometer at Narrabri Observatory I. A. description of the instrument and the observational procedure. 137/, p. 375, Jan. 1967.

[HDA74] Hanbury Brown, R.; Davis, J.; Allen, L. R.: The Angular Diameters of 32 Stars. 167/, pp. 121–136, Apr. 1974.

[He16] He, K. et al.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Pp. 770–778, 2016.

[Hi12] Hinton, G. et al.: Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups. IEEE Signal Processing Magazine 29/6, pp. 82–97, 2012.

[HT56] Hanbury Brown, R.; Twiss, R. Q.: Correlation between Photons in two Coherent Beams of Light. Nature 177/4497, pp. 27–29, Jan. 1956.

[La09] Labeyrie, A. et al.: Luciola hypertelescope space observatory: versatile, upgradable high-resolution imaging, from stars to deep-field cosmology. Experimental Astronomy 23/1, 2009, URL: https://doi.org/10.1007/s10686-008-9123-8.

[LBH15] LeCun, Y.; Bengio, Y.; Hinton, G.: Deep learning. Nature 521/7553, pp. 436–444, May 2015.

[PP18] Pasquet-Itam, J.; Pasquet, J.: Deep learning approach for classifying, detecting and predicting photometric redshifts of quasars in the Sloan Digital Sky Survey stripe 82. A&A 611/, A97, 2018.

[Su12] Sutskever, I. et al.: ImageNet Classification with Deep Convolutional Neural Networks. In: Advances in Neural Information Processing Systems. Vol. 25, Curran Associates, Inc., 2012.

[Su14] Sutskever, I. et al.: Sequence to Sequence Learning with Neural Networks. In: Advances in Neural Information Processing Systems. Vol. 27, Curran Associates, Inc., 2014.

[Zm20] Zmija, A. et al.: LED as laboratory test source for astronomical intensity interferometry. Opt. Express 28/4, pp. 5248–5256, Feb. 2020.

[Zm21] Zmija, A. et al.: Optical intensity interferometry lab tests in preparation of stellar diameter measurements at IACTs at GHz photon rates. Monthly Notices of the Royal Astronomical Society 509/3, pp. 3113–3118, Oct. 2021.

Recht und Technik: Datenschutz im Diskurs (RuT)

Recht und Technik – Datenschutz im Diskurs

Rüdiger Grimm[1], Gerrit Hornung[2], Christoph Sorge[3], Indra Spiecker genannt Döhmann[4]

Vorwort zu den Workshopbeiträgen

Der Workshop „Recht und Technik – Datenschutz im Diskurs“ ist nach neun Jahren ein kontinuierlicher und fester Bestandteil der Jahrestagung der Gesellschaft für Informatik. In den beiden letzten Jahren hatte er pandemiebedingt online stattgefunden. Zu unserer Freude waren es trotz der Online-Distanz sehr lebendige Workshops. Im Übrigen gehörten sie zu den bestbesuchten Workshops der Jahrestagungen.

Dieses Mail wird der Workshop zu unserer großen Freude wieder in Präsenz stattfinden. Inhaltlich bleibt die Ausrichtung wie gewohnt: Wir bieten ein Forum für Beiträge von Informatiker/innen, Juristen/innen und Vertreter/innen benachbarter Fächer, die an Fragestellungen des technikbasierten Datenschutzes arbeiten. Wie schon in den Vorjahren werden Themen adressiert, die anwendungsorientiertes Potential für interdisziplinären Diskurs und Zusammenarbeit bieten und die Möglichkeiten aufzeigen, wie Datenschutz durch Technik präzisiert und umgesetzt werden kann.

Die Beiträge, die wir aus der Vielzahl an qualitätsvollen Einreichungen in einem *peer-reviewed*-Verfahren mit Unterstützung unserer Gutachter/innen auswählen konnten, zeigen die thematische und disziplinären Bandbreite der derzeit im Spannungsfeld von Recht und Technik diskutierten Themen. Die Einreichungen ordnen sich zu folgenden drei Themenfeldern:

- Modellierung und Anwendung von Datensicherheit
- Sicherheit und Datenschutz von E-Mail
- Chancen und Risiken von Online-Verfahren

Wir wünschen den Beiträgen eine rege Diskussion auf dem Workshop und eine interessierte Leserschaft. Wir freuen uns darauf, auch im nächsten Jahr wiederum „Recht und Technik – Datenschutz im Diskurs“ anzubieten.

Unser besonderer Dank gilt den Mitgliedern des Programmkomitees, unter denen in diesem Jahr folgende Personen in die Begutachtung eingebunden waren:

- Matthias Bäcker, Universität Mainz
- Franziska Boehm, KIT

[1] Fraunhofer SIT Darmstadt und Universität Koblenz-Landau <grimm@uni-koblenz.de>
[2] Universität Kassel <Gerrit Hornung <gerrit.hornung@uni-kassel.de>
[3] Universität des Saarlandes <christoph.sorge@uni-saarland.de>
[4] Goethe-Universität Frankfurt am Main <spiecker@jur.unifrankfurt.de>

doi:10.18420/inf2022_43

- Jens-Matthias Bohli, Hochschule Mannheim
- Katharina Bräunlich, Diez
- Matthias Enzmann, Fraunhofer SIT, Darmstadt
- Christian Geminn, Universität Kassel
- Nils Gruschka, Universität Oslo
- Christoph Gusy, Universität Bielefeld
- Marit Hansen, ULD Schleswig-Holstein
- Niko Härting, Rechtsanwalt, Berlin
- Walter Hötzendorfer, Research Institute Digital Human Rights Center, Wien
- Thomas Kahler, DPOblog.eu
- Thomas Kehr, Kanzlei MMV - Martini, Mogg, Vogt PartGmbB
- Thomas Lapp, Rechtsanw, Mediator, Fachanwalt IT-Recht, IT-Kanzlei dr-lapp.de GbR
- Kai von Lewinski, Uni-Passau
- Ronald Petrlic, TH Nürnberg
- Alexander Roßnagel, Universität Kassel und Landesdatenschutzbeauftragter Hessen
- Hannah Ruschemeier, CAIS - Center for Advanced Internet Studies, Bochum
- Burkhard Schafer, Universität Edinburgh
- Tobias Singelnstein, Ruhr-Universität Bochum
- Jürgen Taeger, Universität Oldenburg
- Frederike Zufall, Max-Planck-Institut z. Erforschung v. Gemeinschaftsgütern, Bonn

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 521

A Model-Based Framework for Simplified Collaboration of Legal and Software Experts in Data Protection Assessments

Nicolas Boltz [1], Leonie Sterz [2], Christopher Gerking [3], Oliver Raabe [4]

Abstract: The protection of personal data has become an increasingly important issue. Legal norms focused on data protection, such as the EU General Data Protection Regulation (GDPR), provide legally binding requirements for systems that process personal data. Article 25 of the GDPR refers to the obligation to Data Protection by Design and Default. This can be achieved by conducting legal assessments of the system in the early stages of development and implementing data protection concepts where necessary. This ties in with Article 35, which refers to an obligation to conduct legal assessments before the actual processing of data.

To aid in conducting continuous legal assessments during the design time of software systems, we propose a model-based collaboration framework. This framework not only aids in providing consistent views of the software system for legal experts and software architects but also simplifies communication between both parties. We discuss the overall goals and benefits of such a framework and go into detail about the processes that interact as part of the framework. We also try to align legal concepts with the processes and describe the continuous iterative development using the collaboration framework.

Keywords: data protection by design; legal assessment; GDPR; software architecture; metamodel; design time

1 Motivation

Due to the increasing digital transformation of everyday life, the protection of personal data has become a more and more important issue for society. Legal norms such as the EU General Data Protection Regulation (GDPR) are being increasingly enforced and penalties for non-compliance are high. Also, Article 25 of the GDPR refers to the obligation to Data Protection by Design (DPbD) and Default, i.e. the implementation of data protection concepts during the design time of the system.

[1] Karlsruhe Institute of Technology, Institute of Information Security and Dependability (KASTEL), Dependability of Software-intensive Systems Group (DSiS), nicolas.boltz@kit.edu
[2] Karlsruhe Institute of Technology, Institute of Information Security and Dependability (KASTEL), Information Law for Technical Systems and Legal Informatics Group (ITR), leonie.sterz@kit.edu
[3] Karlsruhe Institute of Technology, Institute of Information Security and Dependability (KASTEL), Dependability of Software-intensive Systems Group (DSiS), christopher.gerking@kit.edu
[4] Karlsruhe Institute of Technology, Institut für Informations- und Wirtschaftsrecht (IIWR), Information Law for Technical Systems and Legal Informatics Group (ITR), raabe@kit.edu

doi:10.18420/inf2022_44

One way to achieve DPbD is to offer concepts during the design time of a system, that aid in the process of designing a system focused on preserving data privacy. During design time, software architecture models are created, which represent the structure and context of the system under development (SUD). Using only software architecture models, it is already possible to derive quality attributes of the future system such as confidentiality [Se22] (besides other attributes like performance or reliability [Re16]). Through an iterative development process, quality issues can be found by analyzing the architecture models, and later resolved by fixing or optimizing the system's architecture. As the cost of fixing an issue increases drastically the later it is fixed [BMU75], in the best case issues should be fixed during design time. Consequently, data protection issues should also be addressed during the design of the system. Furthermore, besides the obvious privacy problems for the individual, which may arise from the improper processing of personal data, issues with data protection potentially entail high penalties and loss of reputation for the organizations handling the data.

While there already exist approaches for privacy and data protection engineering during design time [Ah18; Bi16; Si19] and the definition of context-specific legal requirements [Gr20; MCP21; To21], the interdisciplinary communication and collaboration between legal and technical domains is often overlooked. However, this subject, in particular, has the potential to be critical, as both professions are highly specialized in their field, and in-depth knowledge about the other profession can not be assumed. Even more critical is the fact that similar-sounding terms regarding privacy and data protection might refer to disparate underlying understandings or concepts. During the collaboration, e.g. while conducting a legal assessment of a software system, these similarities, which constitute the basis of the conversation, create the need to first define a uniform terminology to avoid potential misunderstandings. Even if a uniform terminology can once be successfully established, new communication problems are likely to emerge in case of legal change. If the legal situation or its interpretation change over time, these changes must be communicated to the technical personnel, so that they can be reflected in the design of the SUD. A failure to communicate such changes quickly and correctly can not only delay the rollout of systems. If unnoticed, it may even lead to the rollout of non-compliant systems involving the aforementioned financial penalties or cost-intensive product recalls.

We propose a collaboration framework, where direct communication between legal experts and technical personnel is simplified. The framework aims to aid in DPbD and the development of legally compliant systems. Software architects plan a software system, creating a software architecture model in the process. This software architecture model represents a technical view of the system. Information about the software system, required by the legal experts, is provided by another specialized view that is created by transferring, translating, and summarizing information from a software architecture model. This view omits technical information not relevant for a legal assessment and visualizes the system in a way that is more catered toward the legal domain. Furthermore, legal experts use this view to perform a subsumption, identifying a system-specific set of rules, which cover more

complex legal norms. This set of rules is translated to technical analysis queries, that are used to analyze the actual software architecture model. If violations of data protection law are found using the translated rules, the software architecture model can be adjusted as consequence. Each view of the system is model-based and relationships between models are based on transformations or manual processes. As the iteration of the framework ends in the potential adjustment of the planned software system, the resulting software architecture model can feed into a new iteration. This allows for the *continuous* development of the software architecture and conducting of legal assessments at the same time.

The remainder of this paper is structured as follows: Section 2 gives an overview of the foundations regarding the topic of software architecture modeling and data protection legal assessments. Section 3 describes the problems we aim to address and the goals we aim to achieve with the proposed framework. In Section 4 we describe the proposed collaboration framework in detail. In Section 5 concepts from legal methodology are mapped to the corresponding framework element. Section 6 describes the continuous development and assessment process. Section 7 presents related work and Section 8 concludes the paper.

2 Foundations

In this section, we give a brief overview of the main concepts that make up our proposed collaboration framework. In Subsection 2.1 we describe two established approaches to modelling and analyzing software architecture. Subsection 2.2 provides an overview of what a legal assessment is and the different guidelines and methodologies that exist.

2.1 Modelling and Analyzing Software Architecture

There are multiple approaches to describe and analyze software architecture. We focus on Data Flow Diagrams (DFDs) as a lightweight data-centric representation of a software system and the Palladio Component Model (PCM) [Re16] as a detailed representation.

DFDs based on the notion of DeMarco [De79] are unidirectional graphs representing the flow and processing of data in a system. Graph nodes are either Actors, Processes, or Stores. Edges that connect nodes are called data flows and describe a data transmission between the connected nodes. Actors are the source and sink nodes, which start or terminate a sequence of data flows. Process nodes transform incoming data and pass it on as outgoing data. Store nodes hold or emit data. The approach of Seifermann et al. [Se22] extends the DFD notion and provides a way to analyze DFDs for illegal or unwanted data flows. Queries that describe what kind of data flows are unwanted are defined in a domain-specific language (DSL) [Ha21].

Palladio is a tool-supported software architecture simulation approach, that is used to predict an architecture's quality of software properties [BKR09]. The PCM is a model of

component-based software architectures. The basic PCM is made up of several metamodels, each representing a different architectural view of a system [Re16]. The approach of Seifermann et al. [Se22] also provides a way to analyze an extended variant of the PCM using the same DSL as described above.

2.2 Legal Assessments

During a legal assessment, the *facts relevant for the legal assessment*[5] are examined for their legal conformity. Usually, such an assessment refers to an aspect of the law or a specific area of the facts. Our framework has a focus on data protection law. Therefore, our legal assessment concentrates on substantive lawfulness according to Art. 6 GDPR as well as the principles relating to the processing of personal data such as data minimization and purpose limitation (Art. 5 GDPR) and security of processing (Art. 32 GDPR). Article 35 requires that data protection impact assessments should be conducted where the type of processing of personal data results in a high risk to, or if there is a possible change in the risk of the privacy of the natural person. Instead of only focusing on the impact on high-risk areas, our assessment is a more general kind and considers the whole system.

3 Goals

In this section, we describe the problems we aim to address and the goals we aim to achieve with our framework.

The most prominent problem, which we have already touched on in Section 1, is the knowledge gap between legal experts and software architects or engineers. In the legal domain, for example, anonymization is measured using objective factors, such as cost and time for re-identification [GD16, Rec. 26]. However, in the technical domain, anonymization is often measured using concepts from theoretical computer science and mathematical definitions, like k-Anonymity [Sw02]. By providing our framework, we aim to reduce and simplify direct communication that is necessary while conducting a legal assessment, which in turn reduces the potential for misunderstandings. Further, the predefined models and automatic transformations reduce the need to define a uniform terminology, making collaboration more effective and simple.

Another problem that arises due to our view-based approach is maintaining consistency between the views. The software architecture model represents the SUD and is subject to constant changes and development. A view of the SUD catered towards the legal domain, could simply be derived from the software architecture model once at a certain point during development. However, since changes of the software architecture can occur frequently, the legal view might be outdated shortly after. A legal assessment that is based on an outdated

[5] Translated from German legal term ”Sachverhalt”

view might come to false conclusions, which do not correspond to the actual SUD. Even in later phases, the legal view of the system must always be consistent with the software architecture, so that the change to the system can be assessed. Our proposed framework achieves this through the automatic transformation from software architecture model to legal view.

The other way is more complex, as the legal view is used to conduct a legal assessment and identify legal consequences. These consequences can entail requirements and potential changes, that legal experts can not sensibly incorporate into the system. However, our framework provides the legal experts with a way to formulate the legal requirements, that allows for automated translation into technical analysis queries. These queries in turn can be used to automatically analyse the software architecture model for compliance with the legal requirements. This way, our proposed framework allows the views of the system to evolve at the same pace and enables the development of software architecture and conducting legal assessments in parallel.

Wright et al. [WFR13] describe, that the assessments of the privacy/data protection should be a process that begins at the earliest possible stages of system development and continue even after deployment. Our proposed framework aims at providing such a process, by enabling continuous legal assessment of the system. If the software architecture model is kept consistent with the actual runtime system, our proposed framework can also be applied in subsequent phases of system development, keeping an already running system compliant with data protection law. Through continuous assessments, changes in the system can be addressed more quickly. This is supported by the fact that an already existing assessment conclusion, in the form of formulated legal requirements, might only need minor adjustments and the technical analysis process of the software architecture can be performed automatically. Changes in the interpretation of data protection law, due to new legal norms or court rulings, can be addressed similarly.

4 Framework Description

As briefly described in Section 1 and Section 3, the proposed framework aims to simplify communication between legal experts and software architects and in turn aid in DPbD and the development of legally compliant systems. As shown in Figure 1, each view of the system is model-based and consists of a structural model representing the SUD and a rule model representing legal rules that are either derived from or applied to the corresponding structural model.

On the technical side, the structural model is a software architecture model, as described in Subsection 2.1. As the software architecture model informs how the SUD is eventually implemented, it undergoes constant changes during planning and development. Information about the software system, required by the legal experts, is provided by the structural model of the legal view. This structural model called *Legal Assessment Facts Model* is part of our

proposed approach. It contains the information necessary for legal assessments of the system. A model is created by transforming relevant information from the software architecture model and shifting the focus in a way that is more intuitive for legal experts. Legal experts enrich the resulting *Legal Assessment Facts Model* with law-specific elements that cannot be derived from the software architecture model or modelled in the software architecture model. Based on the enriched model, properties of the system that are for example defined in the GDPR, such as *applicability* according to Article 2, *purpose limitation* according to Article 5(1)c or *lawfulness of processing* according to Article 6(1), can already be checked automatically. With these checks, elements that do not fulfill these can be identified. Based on the identified elements, legal experts can either implement legal measures, such as adapting consent forms, contracts, or purpose definitions, to establish the properties, or inform the technical side to initiate changes in the software architecture model. Using the *Legal Assessment Facts Model*, the legal experts conduct further legal assessment and identify additional legal consequences, which apply to the system. These legal consequences can require the system to have specific properties or, make active changes to the system. The changes are often not trivial to implement correctly, which is why they cannot simply be implemented by legal experts in the *Legal Assessment Facts Model* and transformed back to the software architecture model. Instead, the legal experts manually derive a *Legal Norm Rules Model*. This model represents the rule model of the legal view. Based on the identified legal consequences, the legal experts derive legal requirement rules, which are translated into technical analysis queries. The queries can subsequently be used to check the software architecture model for compliance with the legal requirements.

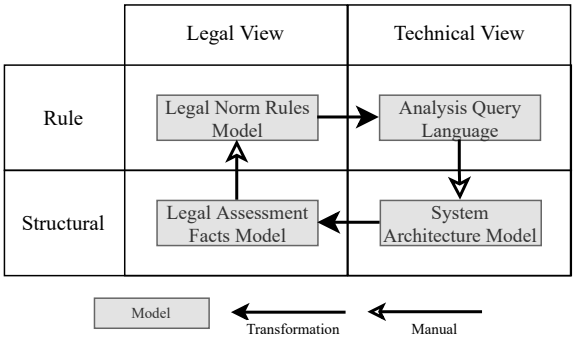


Fig. 1: Models and relationships, categorized by type of information and domain.

5 Legal Methodology

To better align the overall technical process described in Section 4 with the legal domain, we describe the legal concepts that are part of the process. Figure 2 shows the framework with annotated legal concepts. As the software architecture model describes how the SUD is eventually implemented, it represents the *matter of fact*[6]. Based on the matter of fact,

[6] Translated from German legal term ”Tatsache”

the *facts relevant for the legal assessment*[7], need to be captured. As we aim to aid in conducting legal assessments of software systems, the transformation step between software architecture model and *Legal Assessment Facts Model* captures the facts which are relevant for a data protection legal assessment. The structure of the *Legal Assessment Facts Model* is influenced by the information necessary for legal assessments, but also by data protection legal norms in general.

The manual step in which the *Legal Norm Rules Model* is created starts with the identification of the relevant legal norms by the legal expert. The laws contain various *legal consequences*[8] and the respective prerequisites, which are called *definitional elements of a rule*[9]. If all definitional elements of a rule are met, the corresponding legal consequence applies. The legal expert uses the *Legal Assessment Facts Model* to check whether the definitional elements are met. This step is called *subsumption in the narrower sense*[10] in continental European law [Ra12]. The *interpretation of a law*, for example on the basis of court decisions, can also play a role here. Depending on whether or not all relevant facts are fulfilled, it can be determined whether the legal consequence under consideration applies. This is also called *subsumption in the broader sense*[11]. The legal expert transfers the results of his examination into logical and technically readable rules. These result in the *Legal Norm Rules Model*.

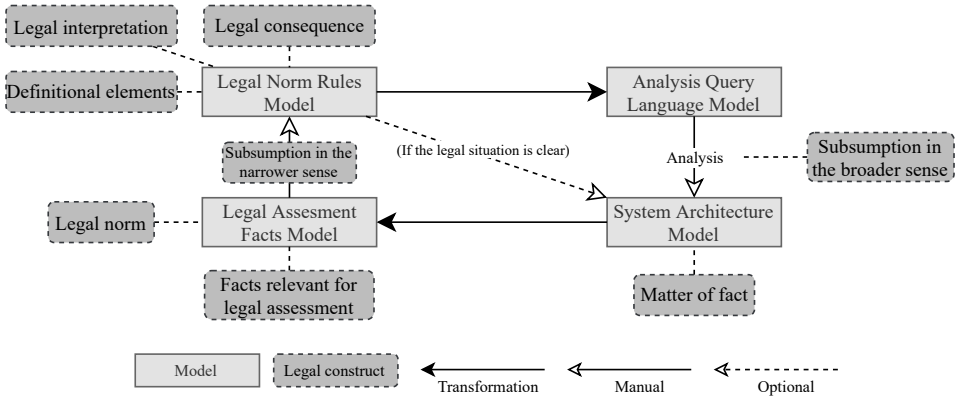


Fig. 2: Concepts of legal methodology, assigned to the corresponding framework elements.

6 Continuous Assessment Process

The process starts as a technical software architecture modeling activity. A software architect plans a software system, creating a software architecture model which represents the SUD.

[7] Translated from German legal term ”Sachverhalt”
[8] Translated from German legal term ”Rechtsfolge”
[9] Translated from German legal term ”Tatbestandsmerkmal”
[10] Translated from German legal term ”Subsumtion im engeren Sinne”
[11] Translated from German legal term ”Subsumtion im weiteren Sinne”

Using a transformation algorithm, information that is relevant for legal assessments is extracted from the software architecture model and a *Legal Assessment Facts Model* is created. A legal expert uses the *Legal Assessment Facts Model* to perform a subsumption in the narrower sense (see Subsection 2.2). As described in Section 4, a *Legal Norm Rules Model* is derived which is custom to the SUD, as represented in the *Legal Assessment Facts Model*. In the following activity, the *Legal Norm Rules Model* is translated to a technical *Analysis Query Language Model*. During this process, trance links from the first transformation to the analysis *Legal Assessment Facts Model* are used to link rules to actual elements of the software architecture model. Using the analysis process of Seifermann et al. [Se22] described in Subsection 2.1, the software architecture model is analyzed with the queries from the resulting *Analysis Query Language Model* to determine if the SUD meets all data protection requirements. Any found data protection requirements that are not met, represent an architectural element or design decision in the software architecture model, that is not compliant with data protection law and requires adjustment. How the software architecture needs to be adjusted is either evident from the requirement or needs to be determined by the software architect.

After the initial iteration, the proposed process is made up of two concurrent loops: The first loop, shown on the right side of Figure 3, only contains technical activities. Once a *Legal Norm Rules Model* has been created by a legal expert, this loop can be iterated independently from any legal cooperation. The software architect can continue to adjust and further plan the software architecture model, analyzing the system's legal conformity using the most current *Legal Norm Rules Model* available. With each iteration, it can be assured that at least a subset of data protection requirements that are not met by the software architecture can be identified. Rules that reference a software architecture element, that has been removed or substantially modified, can be ignored by the analysis. Data protection requirements that are not met, can then be mitigated by making adjustments to the software architecture and starting a new iteration of this loop.

The second loop, on the left side of Figure 3, primarily contains legal activities. Similar to the initial iteration of the process, the now newly adjusted software architecture model is transformed to an updated *Legal Assessment Facts Model*. Based on the updated model, a legal assessment is conducted and the *Legal Norm Rules Model* is updated as a result. The updated *Legal Norm Rules Model* is then translated to a technical *Analysis Query Language Model*, which is used to automatically analyze the software architecture model to find data protection requirements that are not met. If the software architecture is changed, either as a result of a previously not met data protection requirement or as part of the development of the system, a new iteration of this loop is started.

The main benefit of our proposed approach is that both loops can be iterated independently from each other. This allows for the continuous development of the software architecture and conducting of legal assessments at the same time. As each loop ends in the adjustment of the planned software system, the results of a loop feed into the next iteration.

As can be seen in Figure 3, the communication activities that connect technical and legal activities are all based on automated transformations, removing the need for direct interaction. Only if there are very relevant legal consequences, which are obvious and that need immediate attention, direct interaction is necessary.

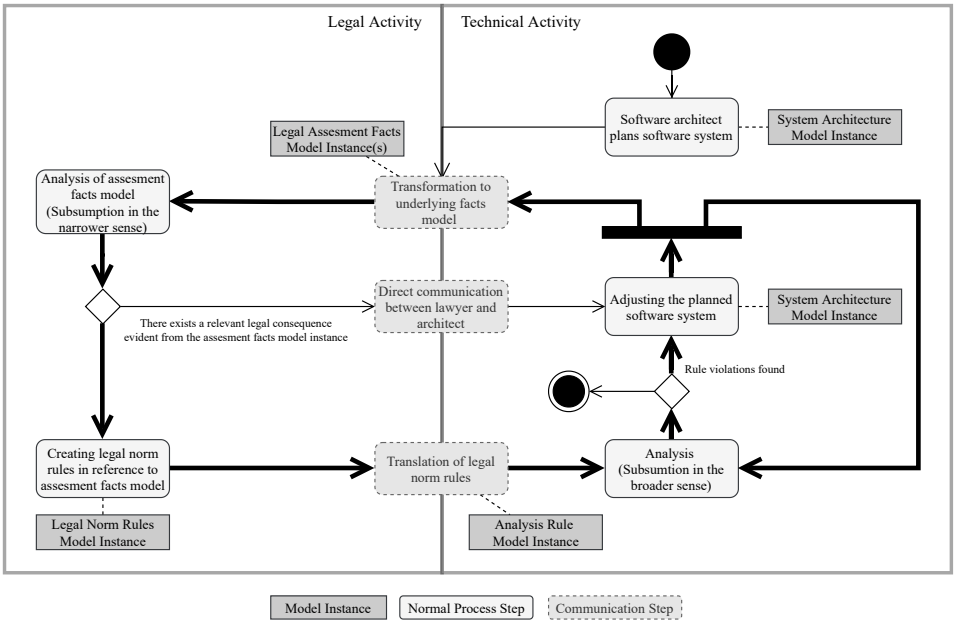


Fig. 3: Activity diagram of the continuous assessment and development process.

7 Related Work

In this section, we discuss other approaches that are related to our proposed framework. Furthermore, as some elements of our described approach still need to be implemented, parts of related work also constitute a base for future work.

In the years since the GDPR was published, there have been a variety of model-based approaches to legal compliance and interdisciplinary cooperation. Torre et al. [To21] define a visual representation of the GDPR, as a way to provide a communication bridge between IT and legal experts. They define a generic GDPR model and invariants expressed in the Object Constraint Language (OCL) that can be used for automated GDPR compliance checking. Further, they describe how the generic model can be specialized depending on the context of the application. However, the specialized model is not derived from a system representation using a formal transformation but rather created by manually changing variation points of the GDPR model. Consequently, changes in the system require manual adjustment of the specialized model. Additionally, changes in the interpretation of law require direct changes

of OCL constraints or the addition of variation points, which might be challenging for legal experts. Sion et al. [Si19] propose an architectural view for DPbD. The view is based on a model, that is derived from definitions of the GDPR. They describe how a transformation could keep the view consistent with a design time model so that there is no discrepancy between view and the actual system. While their approach covers similar aspects as the structural part of our proposed collaboration framework, the process which involves legal experts utilizing the DPbD view and feeding information back into the design time model is not explored. Ahmadian et al. [Ah18] propose a methodology to support data protection impact assessments by performing model-based privacy analyses during design time. They analyze the design of a system and where necessary suggest privacy controls to improve the design. The privacy requirements are defined by annotating a UML diagram with elements of a privacy-focused UML profile. However, the privacy UML profile extension does not directly reference requirements defined by the GDPR. Further, as only a technical UML view of the system is defined, there exists no tailored view for legal experts to work with.

The previously described approaches focus on representing and extending a model in a way that it can be either manually or automatically assessed. In correspondence to our described *Legal Norm Rules Model*, there also exist approaches that focus on the definition of specific legal requirements. Grifo et al. [Gr20] define a domain-specific visual modeling language, which provides for a visual symbolic representation for legal statements. It aims to support legal experts in identifying legal requirements and to assist in the visual representation of case scenarios. Merigoux et al. [MCP21] define the programming language CATALA, which is designed to allow legal experts to translate law into a formal form, that can be further translated into an executable implementation. This approach could be used to represent the *Legal Norm Rules Model*, as it allows for the definition of legal requirements and already provides a translation, which could be leveraged to analyse the SUD.

8 Conclusion and Future Work

In this paper, we present a collaboration framework for continuous legal assessments of software systems during design time. Our proposed framework aims to aid in DPbD and the development of legally compliant systems.

The framework is made up of two views: A technical view, aimed at the technical domain and a legal view aimed at the legal domain. Each view is made up of two models. One represents the structure of the SUD the other a set of rules for compliance with data protection law. The relationships between models of different views are based on automated transformations. Software architects plan a software system, creating a software architecture model. Information about the software system, required by the legal experts, is provided by transferring information from the software architecture model to a *Legal Assessment Facts Model*. The *Legal Assessment Facts Model* omits technical information not relevant for a legal assessment. Legal experts enrich the resulting *Legal Assessment Facts Model* with law-specific elements that cannot be derived from the software architecture model. They

use the *Legal Assessment Facts Model* to apply the law and transform their results into a formal form, creating a *Legal Norm Rules Model* in the process. The *Legal Norm Rules Model* is translated to a technical *Analysis Query Language Model*, which is used to analyze the actual software architecture model.

We define the problems we aim to solve and describe the continuous process, that allows for simultaneous software development and legal assessments. Our framework combines technical and legal methodology and workflows, while at the same time reducing and simplifying the required direct communication between both domains. Through the iterative nature and automated transformations that are part of the proposed process, both views of the SUD are kept consistent with each other. The continuous assessment process not only aligns with the requirements of law but also allows for quicker reactions to changes in the system or law respectively.

In future work, we plan to further implement parts of our proposed framework. Using related work and interdisciplinary collaboration we aim to define a concrete *Legal Assessment Facts Model* and *Legal Norm Rules Model* suited for our described purpose. Further, we plan on evaluating the technical aspects of our proposed framework with a case-study-based evaluation and the applicability and effectiveness with a user study. As this work is part of mobility-focused projects, in a first effort we aim to create a case study for evaluation which is thematically located in the area of connected cars and vehicle to X communication.

Acknowledgments

This publication is partially based on the research project SofDCar (19S21002), which is funded by the German Federal Ministry for Economic Affairs and Climate Action. This work was also supported by funding from the topic Engineering Secure Systems of the Helmholtz Association (HGF) and by KASTEL Security Research Labs (46.23.03).

References

[Ah18] Ahmadian, A. S.; Strüber, D.; Riediger, V.; Jürjens, J.: Supporting privacy impact assessment by model-based privacy analysis. In: Proceedings of the 33rd Annual ACM Symposium on Applied Computing. Pp. 1467–1474, 2018.

[Bi16] Bieker, F.; Friedewald, M.; Hansen, M.; Obersteller, H.; Rost, M.: A process for data protection impact assessment under the european general data protection regulation. In: Annual Privacy Forum. Springer, pp. 21–37, 2016.

[BKR09] Becker, S.; Koziolek, H.; Reussner, R.: The Palladio component model for model-driven performance prediction. Journal of Systems and Software 82/1, pp. 3–22, 2009.

[BMU75] Boehm, B. W.; Mcclean, R. K.; Urfrig, D.: Some experience with automated aids to the design of large-scale reliable software. Transactions on Software Engineering/1, pp. 125–133, 1975.

[De79] DeMarco, T.: Structure analysis and system specification. In: Pioneers and Their Contributions to Software Engineering. Springer, pp. 255–288, 1979.

[GD16] GDPR - The European Parliament and the Council of the European Union: EU General Data Protection Regulation (GDPR), Apr. 27, 2016, URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

[Gr20] Grifo, C.; Teixeira, M. D. G. D. S.; Almeida, J. P. A.; Gailly, F.; Guizzardi, G.: LawV: Towards an ontology-based visual modeling language in the legal domain. In: ONTOBRAS 2020. Vol. 2728, pp. 75–88, 2020.

[Ha21] Hahner, S.; Seifermann, S.; Heinrich, R.; Walter, M.; Bureš, T.; Hnětynka, P.: Modeling data flow constraints for design-time confidentiality analyses. In: International Conference on Software Architecture Companion (ICSA-C). IEEE, pp. 15–21, 2021.

[MCP21] Merigoux, D.; Chataing, N.; Protzenko, J.: Catala: a programming language for the law. Proceedings of the ACM on Programming Languages 5/ICFP, pp. 1–29, 2021.

[Ra12] Raabe, O.; Wacker, R.; Oberle, D.; Baumann, C.; Funk, C.: Subsumtion im engeren Sinne. In: Recht ex machina: Formalisierung des Rechts im Internet der Dienste. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 263–274, 2012.

[Re16] Reussner, R. H.; Becker, S.; Happe, J.; Heinrich, R.; Koziolek, A.; Koziolek, H.; Kramer, M.; Krogmann, K.: Modeling and Simulating Software Architectures – The Palladio Approach. MIT Press, 2016.

[Se22] Seifermann, S.; Heinrich, R.; Werle, D.; Reussner, R.: Detecting violations of access control and information flow policies in data flow diagrams. Journal of Systems and Software 184/, p. 111138, 2022.

[Si19] Sion, L.; Dewitte, P.; Van Landuyt, D.; Wuyts, K.; Emanuilov, I.; Valcke, P.; Joosen, W.: An architectural view for data protection by design. In: International Conference on Software Architecture (ICSA). IEEE, pp. 11–20, 2019.

[Sw02] Sweeney, L.: k-Anonymity: A model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems 10/5, pp. 557–570, 2002.

[To21] Torre, D.; Alferez, M.; Soltana, G.; Sabetzadeh, M.; Briand, L.: Modeling data protection and privacy: application and experience with GDPR. Software and Systems Modeling 20/6, pp. 2071–2087, 2021.

[WFR13] Wright, D.; Finn, R.; Rodrigues, R.: A comparative analysis of privacy impact assessment in six countries. Journal of Contemporary European Research 9/1, 2013.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 533

Datenschutzkonforme Weitergabe von Versichertendaten aus dem Forschungsdatenzentrum

Berna Orak[1], Erik Krempel[2], Arno Appenzeller[3]

Abstract: Die Nachfrage nach der breiten Verfügbarkeit von medizinischen Daten zu Forschungszwecken nimmt stetig zu. Die enormen Datenmengen bieten vor allem für Big Data Verfahren ein großes Potential. In der Gesetzgebung soll dieser Bedarf durch ein Forschungsdatenzentrum, das im Digitale-Versorgung-Gesetz (DVG) geregelt wird, erfüllt werden. Hierbei stellen sich allerdings eine Reihe von Fragen bezüglich des Datenschutzes. So sollen die Daten zwar in pseudonymisierter oder anonymisierter Form vorliegen, allerdings kann nach wie vor ein Re-Identifizierungsrisiko bestehen. Diese Ausarbeitung analysierte die bestehende Gesetzeslage und zieht überblickartig Vergleiche zu internationalen Vorschriften bezüglich der Regulierung anonymer Daten. Auf Basis dieser Analyse wird eine Erweiterung des Forschungsdatenzentrums skizziert, das mithilfe von Privatsphäre-wahrenden Technologien eine datenschutzkonforme Weitergabe von Versichertendaten ermöglichen kann.

Keywords: Forschungsdatenzentrum; Medizinischer Datenschutz; Datenschutzgrundverordnung; Medizinische Daten; Re-Identifizierung; Privatsphäre-wahrenden Technologien; Digitale-Versorgung-Gesetz

1 Einleitung

Das Digitale-Versorgung-Gesetz (DVG)[4] sieht in den §§ 303a-303f SGB V die Etablierung eines Forschungsdatenzentrums vor. Es soll den Zugang von Forscher:innen zu Daten von Millionen Krankenversicherten erleichtern, um unter anderem eine bessere Nutzbarkeit von Gesundheitsdaten für Forschungszwecke zu ermöglichen.[5] Neu ist, dass die Versorgungsdaten von ca. 70 Millionen Bürger:innen zentral gesammelt und der Forschung zur Verfügung gestellt werden. Der Umfang der gesammelten Daten in Verbindung mit ihrer besonderen Vulnerabilität ist deutschlandweit einzigartig und hat in Politik sowie

[1] Goethe-Universität Frankfurt am Main, Lehrstuhl für Öffentliches Recht, Informationsrecht, Umweltrecht und Verwaltungswissenschaft, Institut für Öffentliches Recht, Theodor-W.-Adorno-Platz 4, 60629 Frankfurt am Main, Deutschland, orak@jur.uni-frankfurt.de
[2] Fraunhofer Institut für Optronik, Systemtechnik und Bildverarbeitung (IOSB), Fraunhoferstr. 1, 76131 Karlsruhe, Deutschland, erik.krempel@iosb.fraunhofer.de
[3] Karlsruher Institut für Technologie, Lehrstuhl für Interaktive Echtzeitsysteme, c/o Technologiefabrik Haid-und-Neu-Str. 7,76131 Karlsruhe, Deutschland arno.appenzeller@kit.edu
[4] Gesetz vom 09.12.2019, BGBl. I S.2562.
[5] BT-Drucksache, 19/13438, S. 2.

doi:10.18420/inf2022_45

Wissenschaft Aufmerksamkeit erregt und ebenso Kritik hervorgerufen[6], obwohl die Daten zum Teil anonym zur Verfügung gestellt werden sollen.

Im Rahmen von Forschungsdatenverarbeitung kristallisiert sich längst heraus, dass eine Re-Identifizierung von Daten nicht mehr auszuschließen ist.[7] Gerade im Gesundheitssektor wird die Verarbeitung selbst von anonymen Daten als gefahrvoll betrachtet, da aufgrund der Einzigartigkeit der Angaben technische Mechanismen oftmals keinen allumfassenden Schutz vor einer De-Anonymisierung gewährleisten können.[8] Dabei beschränkt sich die Schwierigkeit nicht nur auf den Gesundheitssektor.[9]. Besonders riskant ist die Weitergabe der Daten an Verarbeiter, die bereits über unterschiedliches Zusatzwissen aus anderen Quellen verfügen. Sie würden zudem bei langfristiger Aufbewahrung von zukünftigen technischen Entwicklungen profitieren, weil sie dann vorhandene Daten genauer auswerten und möglicherweise einzelnen Personen zuordnen könnten.[10] Eine Re-Identifizierungen ist dann nicht ausschließbar.[11]

Behält man diesen Umstand im Blick, stellen sich mit der Etablierung eines Forschungsdatenzentrums, das die Weitergabe der Versorgungsdaten von ca. 90% der Bevölkerung, nämlich der gesetzlich Krankenversicherten, an unbestimmte Nutzungsberechtigte erlaubt, viele Fragen. Wer hat Zugriff auf die Daten? Wie sieht die technische Umsetzung aus? Wie wird verhindert, dass eine solche Re-Identifikation doch möglich ist? Das nach den §§ 303a-303f Sozialgesetzbuch Fünftes Buch (SGB V) vorgesehene Forschungsdatenzentrum lässt diese und weitere Fragen nach der technischen Umsetzung einer Anonymisierung sowie Pseudonymisierung offen. Auch die DSGVO greift die Anonymisierung nicht ausdrücklich auf. Blickt man in die USA und Japan, bestehen dagegen Rahmenregelungen zur Vermeidung eines Re-Identifikationsrisikos.

Der folgende Beitrag betrachtet technische Ausgestaltungsmöglichkeiten zur datenschutzkonformen Weitergabe von Daten an Dritte mit Bezug zum Gesundheitswesen. Im Anschluss an einen Überblick über die Funktionsweise des Forschungsdatenzentrums erfolgt eine Auseinandersetzung mit den rechtlichen Rahmenbedingungen einer Anonymisierung (2). Als entscheidende Vorfrage werden die Herausforderungen der Datenweitergabe an Dritte aus technischer Sicht erörtert (4), um dann der Frage nach Privatsphäre wahrenden Technologien zur datenschutzkonformen Datenweitergabe nachzugehen (5). Ein Fazit schließt den Beitrag (6).

[6] Siehe [Bu19], [SgDB20], [BSgD20], [We20],[KS20].
[7] [MH20], 3574, [RG21], 487.
[8] Caspar, [SHD19], Art. 89 Rn. 55; [La16], 363.
[9] Siehe dazu [RG21], 487.
[10] [Ro21], 176.
[11] [Ro21], 176.

2 Die Datenaufbereitung im neuen Forschungsdatenzentrum und ihre Gefahrpotenziale

Das Forschungsdatenzentrum stellt eine Weiterentwicklung der Datenaufbereitungsstelle dar, die seit 2013 Datenanalysen mit Abrechnungsdaten aus der gesetzlichen Krankenversicherung zu Forschungszwecken durchführt.[12] Es werden Angaben zum Alter, Geschlecht, Wohnort und ICD-Codes im Forschungsdatenzentrum gesammelt.[13] Der ICD-Code ist eine standardisierte Codierung von medizinischen Diagnosen, die von jedermann mithilfe einer Tabelle aufgelöst werden kann. Das bedeutet: Unabhängig davon, ob die Daten pseudonymisiert oder anonymisiert vorliegen, der Datensatz enthält zunächst hochsensible Angaben.

Vorgesehen ist, dass die Datenbereitstellung von drei öffentlichen Stellen des Bundes übernommen wird, vgl. § 303a Abs. 1 SGB V. Der Spitzenverband Bund der Krankenkassen wirkt als sog. Datensammelstelle und das Robert-Koch-Institut (RKI) als sog. Vertrauensstelle. Die Aufgaben des Forschungsdatenzentrums nimmt das Bundesinstitut für Arzneimittel und Medizinprodukte wahr.[14] Die gesetzlichen Krankenkassen geben an die Datensammelstelle, die Angaben für jeden Versicherten mit einem Lieferpseudonym weiter, vgl. § 303b SGB V. Nach Überprüfung auf ihre Vollständigkeit werden die Daten mit einer Arbeitsnummer gekennzeichnet und ohne Lieferpseudonym an das Forschungsdatenzentrum weitergegeben. Das Lieferpseudonym und die Arbeitsnummer erhält nur die Vertrauensstelle, also das RKI. Das Forschungsdatenzentrum übernimmt sodann, wenn Daten zur Forschung angefragt werden, die Aufbereitung der beantragten Daten sowie die Qualitätssicherung und leitet sie an die nutzungsberechtigen Stellen nach Überprüfung der Anträge weiter, vgl. § 303d Abs. 1 SGB V. Die Nutzungsberechtigen erhalten die Daten grundsätzlich in anonymisierter und aggregierter Form.[15] Darüber hinaus prüft das Forschungsdatenzentrum ein spezifisches Re-Identifikationsrisiko beim jeweiligen Nutzungsberechtigen.[16] Die Anforderungen an die technische Umsetzung einer Minimierung des Re-Identifikationsrisikos bleiben dem Forschungsdatenzentrum bzw. dem jeweiligen Auftragsverarbeiter überlassen.[17] Es besteht immerhin die Möglichkeit, die Datenbereitstellung mit der Auflage zu verbinden, eine Zusammenführung der Daten mit externen Datenbeständen zu unterlassen.[18] Insgesamt entspricht die Aufgabenwahrnehmung des Forschungsdatenzentrums somit der Art nach einem Datentreuhänder[19]. Eine allgemeingültige Definition eines Datentreuhandmodells

[12] Michels, [BK22], § 303a Rn. 1a.
[13] Siehe zum vollzähligen Umfang der weitergegebenen Daten: § 303b Abs. 1 i.V.m. § 3 Abs. 1 DaTraV.
[14] Siehe auch § 1 DaTraV.
[15] In Ausnahmefällen können die Daten auch in pseudonymisierter Form zur Verfügung gestellt werden, siehe § 303e Abs. 4 Satz 1 SGB V.
[16] Vgl. § 303d Abs. 1 Nr. 5 SGB V.
[17] Siehe § 303b Abs. 1 Satz 2 und § 303b Abs. 3 Satz 3.
[18] Siehe § 8 Abs. 3 DaTraV.
[19] Siehe [Bu21], 806 ff.

gibt es nicht.[20] Grundsätzlich stellt ein Datentreuhänder einen neutralen Vermittler zwischen Datenverarbeiter und betroffener Person dar, der keinen Mehrwert aus den Daten schöpft. [21]

Die Schwierigkeit der Kontrolle über die Daten beginnt mit der Weitergabe der Daten an die Nutzungsberechtigten. Sie ist gesteigert, wenn sie mit anderen Datensätzen zusammengeführt werden. Die gespeicherten Gesundheitsdaten können z.B. mithilfe von Big Data Analysen ausgewertet werden. Charakteristisch für Big Data Anwendungen ist, dass große Datenmengen (Volume), unterschiedliche Datentypen- und quellen (Variety) sowie eine hohe Geschwindigkeit der Generierung und Verarbeitung der Daten (Velocity) vorliegen.[22] Durch die massenhafte Auswertung unterschiedlicher Daten und Datenquellen können Verknüpfungen hergestellt werden, die wiederum die Identität des Datensubjekts enthüllen oder es jedenfalls identifizierbar machen.[23] Dagegen war die Verarbeitung anonymer Daten in der Vergangenheit eine gut beherrschte Technik, um Identifizierungen zu vermeiden.[24] Dies folgt auch daraus, dass vor wenigen Jahren die Verknüpfung von Datenbanken nur mit unverhältnismäßigem Aufwand möglich war. Mit der Einführung eines Forschungsdatenzentrums werden nun große Datenmengen mit hoher Qualität und Aussagekraft zugänglich.

Zum aktuellen Zeitpunkt befindet sich das Forschungsdatenzentrum noch im Aufbau. Es werden die rechtlichen, technischen personellen und organisatorischen Maßnahmen des neuen Forschungsdatenzentrums definiert und implementiert.[25]. Daneben ist vor dem Bundesverfassungsgericht eine Verfassungsbeschwerde gegen Teile des Digitale-Versorgung-Gesetzes anhängig, so dass noch nicht klar ist, inwieweit es tatsächlich eingeführt wird. Eine einstweilige Anordnung gegen das Gesetz mit dem Ziel Teile des DVG bis zum Hauptsacheverfahren außer Kraft zu setzen, war allerdings erfolglos, da nach Auffassung des Bundesverfassungsgerichts mit der Außerkraftsetzung nachteilige Folgen für wichtige Gemeinwohlbelange verhindert würden (u.a. die medizinische Forschung).[26] Die Vorwürfe gegen das Gesetz sind aber nicht haltlos. Die Diskussion ist geprägt vom fehlenden Widerspruchsrecht der Bürger:innen gegen die Verarbeitung der Daten[27], von der Streubreite und Massenhaftigkeit des Eingriffs bis hin zu europarechtlichen Problemen.[28] Aber eben auch und besonders ein potenzielles Re-Identifikationsrisiko durch die Verknüpfung mit eigenen

[20] Siehe hierzu [K1], 784 f.; [Bu21], 807.
[21] [K1], 784 f.; [Bu21], 807 f.
[22] [De17], 54.
[23] [RG21], 487.
[24] Schaar, [St17], S. 144.
[25] Bundesinstitut für Arzneimittel und Medizinprodukte, https://www.forschungsdatenzentrum-gesundheit.de/das-fdz (zuletzt aufgerufen am 08.07.2022)
[26] BVerfG, Beschluss vom 19. März 2020- 1 BvQ 1/20; siehe hierzu auch: [SgDB20]; von Dewitz, [Ro18], § 303a, Rn. 2.
[27] [We20], 542 f.; anders: Bitkom, Pressemitteilung v. 07.Novemeber 2019, https://www.bitkom.org/Presse/Presseinformation/Bitkom-zum-Digitale-Versorgung-Gesetz (zuletzt aufgerufen am 11.04.2022).
[28] Siehe [SgDB20]; [BSgD20], 990; zum Ausschluss der Betroffenenrechte siehe: [We20], 542 f.; zur Abgrenzung von pseudonymisierten und anonymisierten Daten: [KS20], 43 ff.

Datenbanken der Nutzungsberechtigen ist trotz anonymisierter Datensätze problematisch und hat u.a. den Beschwerdeführer[29] zu seiner Verfassungsbeschwerde bewegt.

3 Rechtliches Fundament anonymer Daten

Anknüpfend an den bereits dargelegten Aufbau des Forschungsdatenzentrums nach dem DVG werden nun die rechtlichen Anforderungen an eine Anonymisierung nach der Datenschutz-Grundverordnung[30] (DSGVO) untersucht sowie ein Überblick über den Umgang Japans und der USA mit anonymen Daten und dem Re-Anonymisierungsrisiko gegeben.

3.1 Art. 9 Abs. 1 und Abs. 2 DSGVO

Art. 4 Nr. 15 DSGVO[31] legaldefiniert als besondere Daten „personenbezogene Daten, die sich auf die körperliche oder geistige Gesundheit einer natürlichen Person, einschließlich der Erbringung von Gesundheitsdienstleistungen, beziehen und aus denen Informationen über deren Gesundheitszustand hervorgehen". Die im Forschungsdatenzentrum gespeicherten Daten stellen solche Gesundheitsdaten dar. Die Verarbeitung von Gesundheitsdaten ist grundsätzlich verboten, vgl. Art. 9 Abs. 1 DSGVO, und nur ausnahmsweise erlaubt. Sie geht mit erheblichen Gefahren für die Grundfreiheiten und Grundrechte der betroffenen Person einher.[32] Der Gesetzgeber hat daher in Art. 9 Abs. 2 DSGVO besondere Voraussetzungen für die ausnahmsweise Verarbeitung von sensiblen Daten geschaffen. Diese Öffnungsklauseln erlauben den Mitgliedstaaten weitergehende mitgliedstaatliche Regelungen zu erlassen.[33] Neben der Einhaltung der allgemeinen Prinzipien aus Art. 5 DSGVO, wie beispielsweise die Zweckbindung, Datensparsamkeit und Transparenz sowie einer Rechtsgrundlage, Art. 6 Abs. 1 DSGVO[34], stellt Art. 9 Abs. 2 DSGVO darüberhinausgehende spezifische Anforderungen für die Verarbeitung von Gesundheitsdaten auf.

Die Weitergabe der Versorgungsdaten durch die Krankenkassen an das Forschungsdatenzentrum auf der Basis von §§ 303a-f SGB V wird durch die Öffnungsklauseln des Art. 9 Abs. 2 lit. h und lit. j) DSGVO ermöglicht. Dies setzt u.a. zusätzlich voraus, dass die Verarbeitung für wissenschaftliche Forschungszwecke angemessene und spezifische Maßnahmen zur Wahrung der Grundrechte der betroffenen Person vorsehen und die Verarbeitung gem. Art. 89 Abs. 1 DSGVO erforderlich ist.

[29] Der Beschwerdeführer leidet an einer seltenen Erbkrankheit und befürchtet trotz Pseudo- oder Anonymisierung aus den Datensätzen re-identifiziert werden zu können.

[30] VO (EU) 2016/679.

[31] Die Kompetenz der EU für das Datenschutzrecht ergibt sich aus Art. 16 Abs. 2 UAbs. 1 S. 1 AEUV, siehe auch Hornung/Spiecker gen. Döhmann, [SHD19], Einleitung, Rn. 155 ff.

[32] Petri, [SHD19], Art. 9, Rn. 1.

[33] Petri, [SHD19], Art. 9, Rn. 101 f.

[34] Zum Verhältnis zwischen Art. 9 und Art. 6 DSGVO: Wedde, [D0], Art. 9, Rn. 3; Petri, [SHD19], Art. 9, Rn. 3; Weichert, [KB20], Art. 9 Rn. 4,; aA: Schantz, [SW17], Rn. 705.

3.2 Art. 89 DSGVO

Art. 89 stellt keinen Erlaubnistatbestand dar, sondern knüpft die an sich zulässige Verarbeitung nach Art. 6 DSGVO und Art. 9 DSGVO nochmals an bestimmte Anforderungen.[35] Damit bezweckt die Vorschrift u.a., den Datenminimierungsgrundsatz aus Art. 5 Abs. 1 lit. c DSGVO zu gewährleisten.[36] Darüber hinaus dient die Regelung dazu, einen Ausgleich zwischen der informationellen Selbstbestimmung und den in der DSGVO gewährten Privilegierungen zugunsten der Forschung zur Berücksichtigung der Forschungsfreiheit herzustellen.[37]. Der Vorbehalt geeigneter Garantien für die Rechte und Freiheiten der betroffenen Person erfordert, „dass technische und organisatorische Maßnahmen bestehen, mit denen insbesondere die Achtung des Grundsatzes der Datenminimierung gewährleistet wird“. Die Maßnahmen in Art 89 Abs. 1 DSGVO verfolgen einen abgestuften Ansatz.[38] Sofern die Zwecke der Verarbeitung mittels einer Anonymisierung erreicht werden können, sind die Daten zu anonymisieren.[39] Scheidet dagegen eine Anonymisierung aus, wenn die Zwecke der Verarbeitung sich nur mit personenbezogenen Daten erreichen lassen,[40] ist zunächst eine Pseudonymisierung in Betracht zu ziehen. Erst wenn dies nicht möglich ist, dürfen unmittelbar erkennbare personenbezogene Daten verarbeitet werden.

3.3 Anonymisierung und die DSGVO

Die Vorprüfung der Möglichkeit einer Anonymisierung wird in Art. 89 DSGVO, im Gegensatz zur Pseudonymisierung, nicht ausdrücklich erwähnt. Auch sonst wird der Begriff nicht in den Artikeln der DSGVO verwendet. Lediglich in Erwägungsgrund (EG) 26 zur DSGVO finden sich Erläuterungen zur „Anonymisierung“. Unter anonymen Informationen werden nur solche Daten verstanden, „die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen, oder personenbezogene Daten, die in einer Weise anonymisiert worden sind, dass die betroffene Person nicht oder nicht mehr identifiziert werden kann“.[41] Daraus lässt sich schließen, dass anonyme Daten das Gegenteil von personenbezogenen Daten sind.[42] Die Grundsätze des Datenschutzes gelten nicht für anonyme Informationen, vgl. EG 26 S. 5 zur DSGVO. Somit ist der sachliche Anwendungsbereich der DSGVO bei anonymen Daten nicht eröffnet. Die zitierte Definition findet sich aber “nur“ in einem Erwägungsgrund. Diese entfalten keine rechtliche Bindungswirkung.[43] Ihnen kommt nur

[35] Pauly, [PP21], Art. 89, Rn. 1; Buchner/Tinnefeld, [KB20], Art. 89 Rn. 1.
[36] Buchner/Tinnefeld, [KB20], Art. 89 Rn. 19.
[37] [Ro19], 159.
[38] Pauly, [PP21], Art. 89, Rn. 12.
[39] Caspar, in [SHD19], Art. 89, Rn. 52.
[40] Pauly, in: [PP21], Art. 89, Rn. 12.
[41] Siehe EG 26 S. 5 zur DSGVO.
[42] Karg, [SHD19], Art. 4 Nr. 1. Rn. 20; [RG21], 487; [Gi21], 482.
[43] Zur rechtlichen Qualität von Erwägungsgründen von Gemeinschaftsrechtsakten: EuGH, Urt. v. 19.6.2014, Rs. C-345/13, ECLI:EU:C:2014:2013 – Karen Millen Fashions, Rn. 31.

besondere Beachtung bei der Auslegung der Vorschriften der DSGVO zu.[44] Da bisher keine Definition der Anonymisierung in der DSGVO vorzufinden ist, bietet es sich an, die Terminologie nach der alten Rechtsauffassung des BDSG a.F. heranzuziehen und zu analysieren.

Ursprünglich wurde das Anonymisieren als „das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können“ definiert, vgl. § 3 Abs. 6 BDSG a.F. Bei der Beurteilung, welche Kriterien für den Begriff anonymer Daten im heutigen Recht heranzuziehen sind, kann die nationale Rechtslage a.F. allerdings nur begrenzt Orientierung bieten. Ein direkter Rückgriff wäre ohnehin unionsrechtswidrig.[45] Es sind nicht mehr nur der „unverhältnismäßig hohe Aufwand an Zeit, Kosten und Arbeitskraft“ zu berücksichtigen, sondern auch andere Faktoren, die für eine Identifizierbarkeit sprechen.[46]

Nach heutigem Verständnis der DSGVO liegen anonymisierte Daten vor, wenn kein Personenbezug möglich ist.[47] Dabei muss mit hinreichender Wahrscheinlichkeit feststehen, dass eine Re-Identifizierung ausgeschlossen ist. [48] Das Entfernen des Namens, der Anschrift und des Geburtsdatums genügen hierfür nicht.[49] Die Bewertung der Wahrscheinlichkeit des Risikoeintritts erfordert auch, dass der der Verantwortliche einen Personenbezug durch Dritte zu vertreten hat, selbst wenn er aus ex-ante Perspektive diesen nicht herstellen konnte.[50] Nach EG 26 „sollten alle objektiven Faktoren, wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand, herangezogen werden, wobei die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen sind.“ Der EuGH hat dies präzisiert, dass es vorrangig auf die Mittel ankommt, die legal herangezogen werden können.[51] Damit wird zum Ausdruck gebracht, dass das Datenschutzrecht kein erlaubtes Risiko zulässt.[52] Wie unterschiedlich die Anforderungen über Zeitablauf werden können, zeigt sich schon daran, dass mit den heutigen technischen Möglichkeiten und mit dem Abgleich von Datenbanken an das Merkmal ”unverhältnismäßig hoher Aufwand” durchaus andere Anforderungen zu stellen sind als noch zu BDSG a.F.-Zeiten.[53] Erforderte die Verknüpfung mit anderen Datenbanken früher noch einen hohen zeitlichen Aufwand, ist das mit den heutigen technischen Möglichkeiten in wenigen Sekunden erledigt.

Festzuhalten bleibt, dass die Definition anonymer Daten ein changierender ist, der unter

[44] Wegener, [CR21], Art. 19 Rn. 32.
[45] [Me21], der auch die landesrechtlichen Legaldefinitionen in § 3 BbgDSG, § 4 DSG NRW und § 28 Abs. 3 ThürDSG für unionswidrig einstuft; dagegen: [RG21], 487 , die die Definition weiterhin aufgreifen.
[46] Hierfür spricht der Wortlaut ”wie” im EG 26 S. 4 zur DSGVO; siehe: [Me21]; Ernst, [PP21], Art. 4 Rn. 50.
[47] [Ro19], 159; [MH20], 3574.
[48] [MH20], 3574.
[49] [MH20], 3574.
[50] Ernst [PP21], Art. 4 Rn. 13.
[51] EuGH, Urt. v. 19.10.2016, Rs. C-582/14, – Breyer.
[52] [MH20], 3574; Ernst, [PP21], Art. 4 Rn. 13.
[53] Ernst, [PP21], Art. 4 Rn. 50.

dem technischen Wandel stetig einer Neujustierung bedarf. Dies ruft für den Umgang mit anonymen Daten nach einer Rahmenregelung, die bislang noch nicht existiert. Umso erfreulicher ist das Vorhaben der Bundesregierung, Standards für die Anonymisierung und die Sanktionierung einer De-Anonymisierung etablieren zu wollen.[54] Ob die Etablierung dieser Regelungen ausschließlich auf nationaler Ebene erfolgt oder eine europäische Lösung bestrebt wird, bleibt unklar. Angesichts der europaweiten Wirkung der DSGVO wäre eine europäische Regulierung des Umgangs mit anonymisierten Daten wünschenswert, auch im Hinblick auf den gerade entstehenden Datenmarkt in der EU durch Gesetzeswerke wie z.B. den Data Governance Act, jedenfalls aber für besonders sensible Daten wie anonymisierte Gesundheitsdaten, die auf supranationaler Ebene im Europäischen Gesundheitsdatenraum ausgetauscht werden sollen.[55]

3.4 Überblick USA und Japan

In den USA existiert mit dem überarbeiteten Health Insurance Portabilitity and Accountability Act (HIPAA)[56] seit 2013 eine Regelung für die standardisierte Übermittlung von elektronischen Gesundheitsdaten, die zumindest die Gefahr einer Re-Identifikation berücksichtigt.[57] Der HIPAA listet 18 Merkmale (Name, ortsbezogene Angaben, Sozialversicherungsnummer etc.), die zur Umsetzung einer De-Identifikation entfernt werden müssen.

In Japan dagegen ist der Gesetzgeber einen Schritt weitergegangen und hat die Definition anonymer Daten in den Amended Act on the Protection of Personal Information (APPI)[58] aufgenommen, vgl. Art. 2 Abs. 9 APPI. Im Zuge der Reform des APPI wurden auch Standardisierungsprozesse zu Anonymisierungstechniken, Anforderungen an die Datensicherheit, ein Verbot der Zusammenführung unterschiedlicher Datenquellen sowie Informationspflichten gegenüber der Öffentlichkeit eingeführt.[59]

Der aufgezeigte kursorische Überblick illustriert schon auf den ersten Blick Unterschiede in der Regulierung anonymer Daten, die zwischen Europa, Amerika und Asien herrschen. Unabhängig davon, wie effektiv die ergriffenen Maßnahmen anderer Rechtsordnungen sind, gilt es die Regelungslücke für die Anwendung anonymisierter Daten und damit die Nicht-Anwendbarkeit der DSGVO zu schließen. Die Öffnungsklauseln des Art. 9 Abs. 2 DSGVO ermöglichen dies zumindest für Gesundheitsdaten und damit auch für das Forschungsdatenzentrum. Die Herausforderung für die Umsetzung des Forschungsdatenzentrums liegt

[54] Koalitionsvertrag der Bundesregierung, S.17, https://www.tagesschau.de/koalitionsvertrag-147.pdf (zuletzt aufgerufen am 08.07.2022).
[55] Siehe zum Europäischen Gesundheitsdatenraum: COM (2022) 197 final.
[56] Pub. L 104-191, 104th Congress, 110 Stat.1936 v. 21.8.1996; 42 U.S.C. 210ff, 45 C.F.R. part 160 and subparts A und E of part 164.
[57] [Le13], 767.
[58] Gesetz Nr. 57 von 2003; in Kraft getreten am 1.4.2005.
[59] Siehe hierzu: [GLF18], 417f

aber nicht nur am schillernden Anonymisierungsbegriff, sondern auch in der Auswahl der einschlägigen technischen Infrastruktur, um ein minimales Re-Identifikationsrisiko zu gewährleisten.

4 Herausforderungen der Datenweitergabe an Dritte

Aus einer rein datenschutzrechtlichen Betrachtung wäre es für den Schutz der Rechte und Freiheiten der von der Datenübertragung betroffenen Personen vorteilhaft, wenn diese in einer Form erfolgt, die der Definition von anonymen Daten unter BDSG a. f. entspricht. Das würde bedeuten, dass der Personenbezug von Daten derart aufgehoben ist, dass er nicht oder nur unter unverhältnismäßigem Aufwand an Zeit, Kosten und Arbeitskräften wiederhergestellt werden kann.

Dieser Wunsch ist jedoch in der Realität nur selten erreichbar. Menschen sind nicht nur durch eindeutige Identifikatoren wie Name oder Krankenversichertennummer zu identifizieren. Ebenso lassen sich Personen über eine Kombination uneindeutiger Daten, oft Quasi-Identifikatoren genannt, wie Geschlecht, Geburtsdatums oder die Postleitzahl identifizieren. Besonders eindrücklich demonstrierte dies Latanya Sweeney im Jahr 2000.[60] In ihrer Veröffentlichung hat sie gezeigt, dass sich 87% aller Amerikaner eindeutig aus der Kombination des Geburtsdatums, ihres Geschlechts und ihrer Postleitzahl identifizieren lassen. Da medizinische Versorgungdaten noch deutlich detaillierter sind, muss man davon ausgehen, dass damit ebenfalls eine Re-Identifizierung möglich ist.

Cynthia Dwork geht hier sogar so weit, dass sie sagt: “De-Identified data isn’t”. Damit möchte sie auf die besondere Abhängigkeit zwischen der Anonymisierung von Daten und deren Nützlichkeit eingehen. Wenn garantiert werden soll, dass Angreifer nicht in der Lage sind Personen zu Re-Identifizieren, müssen die Daten sehr stark verändert werden. Das bedeutet, dass alle Identifikatoren entfernt und auch alle Quasi-Identifikatoren so stark verändert werden müssen, dass über diese keine Zuordnung mehr möglich ist. Das kann beispielsweise erreicht werden, in dem das Geburtsdatum auf Monat und Jahr reduziert wird oder in der Postleitzahl die hinteren Stellen entfernt werden. Offensichtlich ist, dass dadurch ein Teil der Aussagekraft der Daten verloren geht. Je mehr die Daten verändert werden, um eine Anonymisierung zu erreichen, desto stärker wird die Nützlichkeit eingeschränkt. Laut Dwork kann nur entschieden werden, ob etwas erzeugt werden soll, was sicher anonymisiert ist oder ob nützliche Daten zurückbehalten werden sollen.

An dieser Stelle wird deutlich, dass der Zweckbindungsgrundsatz in Wechselwirkung mit dem Prinzip der Datensparsamkeit steht[61]. Die Daten dürfen nur für einen bestimmten Zweck verarbeitet werden; gleichzeitig sind sie damit auf das notwendige Maß für diesen Zweck zu beschränken. Der Grundsatz der Datensparsamkeit darf aber nicht dazu führen, dass unbrauchbare Daten erzeugt werden, die für das Forschungsvorhaben letztlich keine

[60] [Sw00]
[61] Siehe zu den Prinzipien der DSGVO Art. 5 Abs. 1 DSGVO.

Aussagekraft haben. Tatsächlich verlangt das Prinzip der Datenminimierung auch nicht nach einer absoluten Beschränkung oder Reduzierung der Datenmenge.[62] Um die Nützlichkeit der Daten zu gewährleisten, muss daher auf die Festlegung eines bestimmten Zwecks geachtet werden. Ist im Ergebnis der Verarbeitungszweck durch anonymisierte Daten erreichbar, würde eine Zuwiderhandlung durch Verarbeitung nicht anonymisierter Daten gegen das Prinzip der Datenminimierung verstoßen.[63] Für die Forschung mit und an Daten stellt sich aber häufiger die Hürde, von Beginn an einen bestimmten Zweck festzulegen.[64] Darüber hinaus steht der Anonymisierung gerade im klinischen Forschungsbereich entgegen, dass eine fortlaufende Zuordnung der Daten bei Langzeitstudien notwendig ist[65].

In der medizinischen Forschung wird oft davon ausgegangen, dass eine echte Anonymisierung bei Erhalt der Nützlichkeit der Daten nicht möglich ist. Gleichzeitig sollen alle möglichen technischen und organisatorischen Maßnahmen ergriffen werden, um die Daten bzw. die betroffenen Personen optimal zu schützen. Die TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. – erarbeitet seit ihrer Gründung im Jahr 2003 Konzepte für eine datenschutzgerechte medizinische Forschung. Ein zentraler Baustein vieler Konzepte ist die Trennung der Daten in die zwei Gruppen IDAT und MDAT. IDAT sind dabei alle Daten wie Namen, Adresse oder Krankenversichertennummer die primär für die Identifikation der Personen benötigt werden. MDAT bezieht sich auf alle Daten die Aussagen über medizinische Sachverhalte wie beispielsweise ein Laborergebnis oder einer Diagnose machen. IDAT werden im Rahmen der Konzepte so verändert, dass ein optimaler Schutz der betroffenen Personen erreicht wird. Beispielsweise indem sie komplett durch ein Pseudonym ersetzt werden und nur noch eine Treuhandstelle eine Zuordnung zwischen Pseudonym und IDAT herstellen kann. MDAT bleiben unverändert und werden beispielsweise dadurch geschützt, dass strenge Zugangs- und Zugriffsregelungen etabliert werden. Das Ziel ist Datensparsamkeit zu gewährleisten, damit Wissenschaftler:innen nur genau die Daten bekommen, die sie für ihre Forschung benötigen. Aufbauend auf den Konzepten der TMF können nun noch weitere Maßnahmen ergriffen werden, um die betroffenen Personen bei einer Weitergabe der Daten optimal zu schützen.

5 Technisches Konzept für ein datenschutzkonformes Forschungsdatenzentrum

Der folgende Abschnitt gibt einen Überblick über verfügbare und nutzbare Privatsphäre wahrende Technologien zur Datenweitergabe. Auf Basis der rechtlichen Analyse aus den vorherigen Abschnitten und den hier aufgezeigten Technologien wird ein Umsetzungsvorschlag zu einem datenschutzkonformen Forschungsdatenzentrum aufgezeigt.

[62] Herbst, [KB20], Art. 5 Rn. 5; [K0] 185.
[63] Herbst, [KB20], Art. 5 Rn. 56.
[64] Roßnagel, [SHD19], Art. 5 Rn. 128; siehe auch: [Sp17].
[65] Weichert, ABIDA Gutachten, 2018, Big Data im Gesundheitsbereich, S. 192, https://www.abida.de/sites/default/files/ABIDA%20Gutachten-Gesundheitsbereich.pdf.

5.1 Privatsphäre wahrende Technologien zur datenschutzkonformen Datenweitergabe

In der Vergangenheit haben sich vor allem für medizinische Studien mit Datenveröffentlichung /-weitergabe Metriken wie k-Anonymität und l-Diversity etabliert.[66] Bei k-Anonymität wird für eine Datenveröffentlichung vorausgesetzt, dass es für jeden Eintrag mit einem als sensitiv erachtetes Attribut (beispielsweise ein medizinischer Befund) $k-1$ andere Datensätze mit demselben sensitiven Attribut existieren.[67] Hierdurch soll das triviale Raten mit Hintergrundwissen über eine seltene Erkrankung verhindert werden. Gruppen mit denselben Attributen werden Äquivalenzklassen genannt. Diese Klassen erhält man durch die Unterdrückung oder Generalisierung von identifizierenden Attributen. Während bei der Unterdrückung das einfache Entfernen beziehungsweise Ersetzen durch einen Platzhalter ausreicht, werden für das Generalisieren Attributs Hierarchien benötigt. So wird zum Beispiel ein spezifisches Alter einer Altersgruppe zugeordnet. Allerdings besitzt k-Anonyimität diverse Schwächen. Sollte es beispielsweise zu einer unabsichtlichen Veröffentlichung der Daten kommen, ist es für die betroffenen Personen irrelevant, dass auch weitere Personen das gleiche sensitive Attribut haben. Um solche Angriffe zu verhindern, wurde die Definition mit l-diversity verschärft.[68] Um die Äquivalenzklassen weiter zu diversifizieren und einen Angriff mit Hintergrundwissen zu erschweren, wird gefordert, dass jede Äquivalenzklasse mindestens l verschiedene sensitive Attribute beinhaltet. Hierdurch würde das einfache Wissen über die Zugehörigkeit eines Individuums zu einer Äquivalenzklasse nicht ausreichen, um das sensitive Attribut zu erhalten. Allerdings gibt es auch eine Reihe effizienter Hintergrundwissens Angriffe auf l-diversity. Beispielsweise die Ähnlichkeitsattacke, bei der die sensitiven Attribute sehr ähnlich oder semantisch synonym sind (beispielsweise verschiedene Arten von Darmkrebs). Neben l-diversity existieren weitere Begriffe wie t-closeness und δ-presence , die aber ebenfalls Angriffsvektoren bieten und weitere nicht immer trivial erfüllbare Anforderungen an die zu weitergebende Daten stellen.[69]

In den letzten Jahren wurde der Begriff der Differential Privacy (DP) nach Dwork ein immer relevanteres Maß für eine Privatsphäre Garantie.[70] Im Gegensatz zu den vorher genannten Techniken, benötigt Differential Privacy weniger semantisches Wissen über die zu privatisierenden Daten. Vielmehr betrachtet DP, wie sehr die Daten eines Individuums das Ergebnis einer Datenauswertung beeinflusst. Gemäß der Definition von Differential Privacy bedeutet dies für zwei Datensätze D_1 und D_2, die sich um genau ein Element unterscheiden, und $S \subseteq Bild(K)$, dass gilt

$$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon) \times \Pr[\mathcal{K}(D_2) \in S].$$

Für eine gegebene Auswertung der Daten, bedeutetet dies, dass für alle möglichen Datensätze das Ergebnis sich höchstens um e^{ϵ} unterscheidet. Um diese Eigenschaft zu erfüllen wird K

[66] [Zu21]
[67] [Sw02]
[68] [Ma06]
[69] [LLV07, NAC07]
[70] [Dw06]

als sogenannter Differential Private Mechanismus genutzt. Mit ihm wird das Ergebnis so stark verändert, bis die Bedingung von DP erfüllt ist. ϵ wird auch das Privatsphärebudget genannt. Ein kleines ϵ bedeutet, dass der Privatsphäreschutz sehr hoch ist. Für diese Fall muss aber ein sehr starker privater Mechanismus angewandt werden, so dass die Nutzbarkeit der Daten sinken kann. Im Umkehrschluss steigt die Nutzbarkeit der Daten bei großem ϵ und die Daten werden weniger stark mit Rauschen belegt. Neben der Wahl des Privaten Mechanismus, spielt noch die Vertrauenswürdigkeit gegenüber der Datenbank eine Rolle. Falls diese vertrauenswürdig ist, kann Differential Privacy in einer globalen Variante verwendet werden, bei der alle Daten vor der Weitergabe zentral verrauscht werden. In diesem Fall kennt die Datenbank die realen Werte. Falls die Datenbank nicht vertrauenswürdig erscheint oder im Szenario der Datenweitergabe an eine dritte nicht vertrauenswürdige Partei, kann die lokale Variante von Differential Privacy, Local Differential Privacy, verwendet werden. Hierbei werden die einzelnen Datenpunkte individuell bei den Datensendenden verrauscht und erst dann weitergegeben. In diesem Fall kennt nur das Individuum die realen Daten.

Neben der theoretischen Erforschung der Differential Privacy Eigenschaft stehen inzwischen auch immer mehr Werkzeuge zur Verfügung die Technologie in der Praxis zu verwenden. Das OpenDP[71] Konsortium, gebildet aus Teilen von Microsoft und der Havard University, stellt zum Beispiel mit SmartNoise eine Programmbibliothek zur Verfügung, mit der sich DP in verschiedenen Anwendungen verwenden lässt. Allerdings löst diese Verfügbarkeit der Technologie noch nicht die diversen offenen Fragen in der Anwendung. So ist umstritten, welches Privatsphäre Budget ϵ verwendet werden soll und die Antwort ist auch Anwendungsfall abhängig. Eine wichtige Frage hierbei ist, wie das Privatsphäreschutz V.S. Nutzbarkeit Problem bewertet werden soll. Ein hoher Privatsphäreschutz durch ein kleines ϵ kann zur Akzeptanz der Technologie beitragen, schwächt allerdings den Forschungsnutzen der Daten. Die Forschungsseite würde vermutlich eher eine bessere Nutzbarkeit der Daten bevorzugen. Darüber hinaus gilt es auch die Differential Privacy Eigenschaft für nichttechnische Nutzende verständlich zu gestalten. Die Angabe eines konkreten Epsilon-Wertes wird vermutlich als eher weniger hilfreich betrachtet, um den Schutz der eigenen Daten bei einer Weitergabe zu bewerten. Hier könnten sich traditioneller Metriken wie k-anonymität oder l-diversity als durchaus leichter erklärbare Variante eignen.

Die zuvor beschriebene problematische Wahl von Epsilon, dessen Auswirkung und die verschiedenen Schwächen der traditionellen Technologien können eine große Unsicherheit für den Einsatz in Real-World Anwendungen bedeuten. Hierfür wäre eine Standardisierung verschiedener Privatsphäre wahrender Technologien denkbar, die den Technologie Einsatz regelt und einen Leitfaden bietet. Beispiele wie der zuvor in Kapitel III erwähnte HIPAA zeigen, dass die standardisierte reine Anonymisierung zwar eine rechtliche Sicherheit bietet, aber keineswegs ein Allheilmittel gegen Re-Identifizierung ist.[72] Im Entwurf für das Forschungsdatenzentrum soll der Versuch der Re-Identifizierung zwar sanktioniert[73] werden,

[71] Siehe: `https://opendp.org` (Letzter Zugriff: 10.03.2022)
[72] Siehe z.B. [JE18])
[73] Siehe § 303 e Abs. 6 SGB V

technische Maßnahmen werden dennoch nicht vorausgesetzt eine solche zu verhindern. Deswegen wäre die Pflicht jegliche geteilten oder veröffentlichten Daten, auf ein Re-Identifizierungsrisiko zu untersuchen und gegebenenfalls dieses Risiko durch den Einsatz zuvor beschrieben Technologien zu mindern, auf jeden Fall ein wichtiger Bestandteil eines solchen Leitfadens.

5.2 Mögliche Umsetzung

Auf Basis des zuvor erläuterten rechtlichen Rahmens und des aktuellen Standes der Technik wird hier eine mögliche Umsetzung für ein datenschutzkonformes Forschungsdatenzentrum aufgezeigt. Der Ansatz verfolgt die Prinzipien von **Transparenz**, **Kontrolle**, **Verfügbarkeit** und **Privatsphäreschutz**.

Abbildung 1 zeigt eine mögliche Architektur für ein datenschutzkonformes Forschungszentrum als Erweiterung des im DVG vorgeschlagenen Forschungsdatenzentrum.

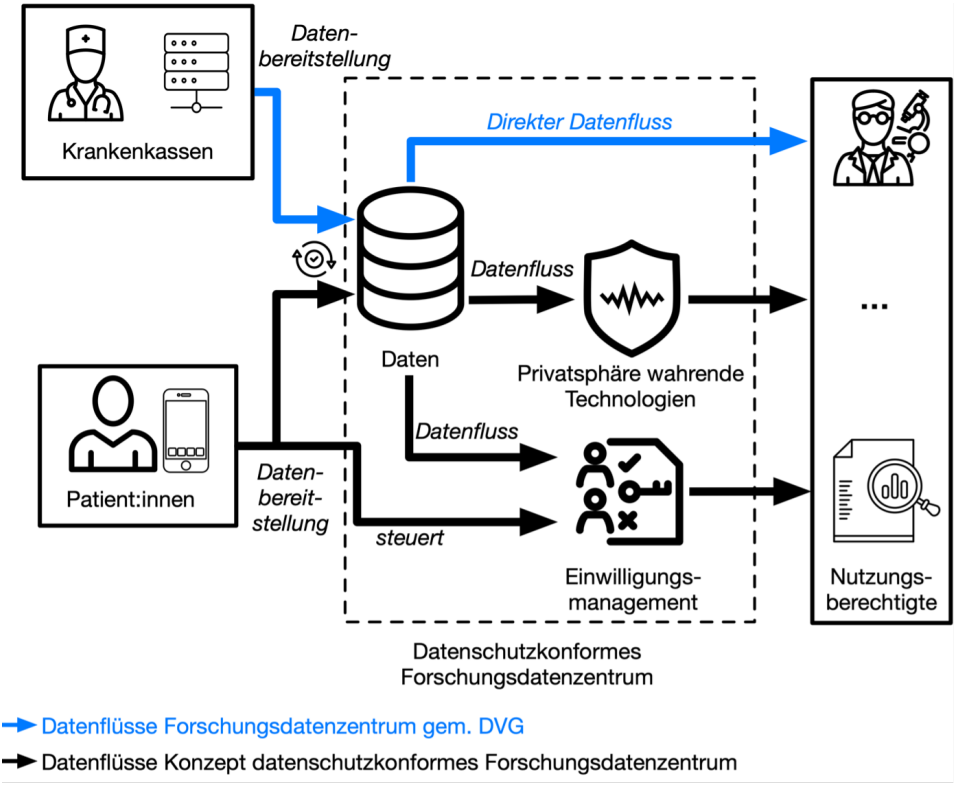


Abb. 1: Technische Architektur datenschutzkonformes Forschungszentrum

Der vorliegende Entwurf sieht im Wesentlichen vier Parteien vor. Zum einen Patient:innen, Krankenkassen, das datenschutzkonforme Forschungsdatenzentrum als Datentreuhänder und Nutzungsberechtigte. Die Daten im Forschungsdatenzentrum werden zum einen von den Datenlieferanten wie Krankenkassen, Ärzten oder Kliniken bereitgestellt, aber können auch beispielsweise direkt vom Patient:innen stammen (wie Lifestylefitnessdaten von Smartwatches). Denkbar wäre auch eine Anbindung an die Telematikinfrastruktur, die für die Elektronische Patientenakte (ePA) genutzt wird. Um Patient:innen Teilhabe zu ermöglichen, sollten Betroffene stets eine Übersicht haben, welche Daten im Forschungszentrum zur Verfügung stehen. Dies ermöglicht ein **transparentes** Verfahren für die Betroffenen. Für die Umsetzung bietet sich eine Smartphone Anwendung an, die auch weitere Verwaltungsmöglichkeiten für medizinische Daten bietet. Über eine solche App könnten Nutzer:innen auch **Kontrolle** über ihre Daten ausüben. So könnte ein Einwilligungsmanagement dazu dienen Daten explizit für Forschungsvorhaben freizugeben. Auf Basis dieser Einwilligungen wäre ein automatisierter Zugriff für die Forschungsinteressierten möglich. Wobei durch diesen Zugriff nur Daten, für die die explizite Einwilligungen von Betroffenen vorliegen, freigegeben werden. Diese Möglichkeit steigert die **Verfügbarkeit** der Daten. Für Massendaten, die beispielsweise in anonymisierter Form aufgrund der verschiedenen Rechtsgrundlagen genutzt werden können, sollten zusätzlich Privatsphäre wahrende Technologien verwendet werden, um auch das Risiko von Re-Identifizierungen zu senken. Somit wären die Forschungsinteressen gewahrt, bei einem gleichzeitig quantifizierbaren **Privatsphäreschutz** für Betroffene.

6 Fazit und Ausblick

Forschung auf bereits vorhandenen Versorgungsdaten in Deutschland hat ein großes Potential für die Verbesserung der medizinischen Behandlung. Im europäischen und internationalen Vergleich liegen die Möglichkeiten in Deutschland weit zurück. Mit dem Digitale-Versorgung-Gesetz wurde ein erster Schritt gegangen, um diese Forschung zu ermöglichen. Leider wurden dabei viele wichtige Fragen des Datenschutzes nicht geklärt. Die aktuell geplante Form der intransparenten Datenweitergaben in der Patient:innen keinerlei Mitspracherecht über die Nutzung ihrer Daten haben, erscheint gleich aus mehreren Gründen ungeeignet. Grundsätzlich stellt sich die Frage, ob dieser Teil des Gesetzes vor dem Bundesverfassungsgericht Bestand haben wird und ob es nicht europarechtswidrig ist. Weiter sehen die Autor:innen die Gefahr, möglicherweise langfristig die Akzeptanz und das Vertrauen der Bürger:innen zu verlieren. Im Vergleich zu dem im Gesetz skizzierten Ansatz, kann die im Paper vorgestellte Architektur die Transparenz und Mitsprache bei Patient:innen stärken. Dies hätte den weiteren Vorteil, dass sich über die Infrastruktur auch weitere Mechanismen, wie etwa eine Datenspende abbilden lassen.

Die aufgezeigten Problemfelder des Forschungsdatenzentrums erfassen nur einen Bruchteil der Fragen, die dessen Etablierung mit sich bringt. Ungeachtet der Transparenz- und Kontrollmöglichkeiten fehlen im aktuellen Stand des Digitale-Versorgung-Gesetzes verbindliche

Vorgaben zum Umgang mit den und damit einem fortgesetzten Schutz der bereitgestellten Daten. Während das japanische und amerikanische Recht sehr konkrete Vorgaben zum Schutz der betroffenen Personen gibt, fehlen diese in der DSGVO und dem DVG. Die Gefahr einer Re-Identifikation zunächst anonymer Daten ist zu befürchten, aufgrund des technischen Fortschritts in der Auswertung von Daten sowie durch die Verknüpfung unterschiedlicher Datensätze. Bisher bietet das geltende Recht allerdings keinerlei Lösung für dieses Defizit. Es verdeutlicht aber, dass technische Entwicklungen für die Gewährleistung des Datenschutzrechts immer wieder überprüft und neubewertet werden müssen.[74]

Mit Differential Privacy existiert eine Technologie, die nicht nur zur De-Identifikation von Daten geeignet ist, sondern gleichzeitig ein Maß für das Risiko einer Re-Identifizierung bietet. Dabei muss jedoch gesagt sein, dass es bisher zu wenig Erfahrung mit der Technologie gibt, um bereits Vorgaben in Form von konkreten Schwellwerten für Differential Privacy zu machen.

Danksagung

Diese Arbeit entstand in enger Zusammenarbeit mit dem KASTEL Projekt.

Literaturverzeichnis

[BK22] Becker, Ulrich; Kingreen, Thorsten: SGB V - Gesetzliche Krankenversicherung. Beck C. H., München, 2022.

[BSgD20] Bretthauer, Sebastian; Spiecker gen. Döhmann, Indra: Das Digitale-Versorgung-Gesetz als Einfallstor für eine Neujustierung von einstweiligem Rechtsschutz vor dem BVerfG und der Eingriffsqualität bei Datenverwendungen. JuristenZeitung, 75(20):990–996, 2020.

[Bu19] Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (BfDI): , Stellungnahme des Bundesbeauftragten für den Datenschutz und die Informationsfreiheit zum Entwurf für das Digitale Versorgung-Gesetz,. https://www.bfdi.bund.de/SharedDocs/Downloads/DE/DokumenteBfDI/Stellungnahmen/2019/StgN_Digitale_Versorgung_Gesetz.pdf?__blob=publicationFile&v=4, 2019. zuletzt aufgerufen am 08.07.2022.

[Bu21] Buchner, Benedikt; Haber, Anna Christine; Hahn, Horst Karl; Prasser, Fabian; Kusch, Harald; Sax, Ulrich; Schmidt, Carsten Oliver: Das Modell der Datentreuhand in der medizinischen Forschung. Datenschutz Datensicherheit - DuD, 45(12):806–810, Dezember 2021.

[CR21] Calliess, Christian; Ruffert, Matthias: EUV/AEUV - Das Verfassungsrecht der Europäischen Union mit Europäischer Grundrechtecharta. Beck C. H., München, 2021.

[De17] Deutscher, Ethikrat: Big Data und Gesundheit : Datensouveränität als informationelle Freiheitsgestaltung : Stellungnahme. Deutscher Ethikrat, Berlin, 2017.

[74] Siehe hierzu: Spiecker gen. Döhmann, in: [Sp20], 489.

[Dw06] Dwork, Cynthia: Differential Privacy. In: Automata, Languages and Programming, Lecture notes in computer science, S. 1–12. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[D0] Däubler, Wolfgang; Wedde, Peter; Weichert, Thilo; Sommer, Imke: EU-DSGVO und BDSG - Kompaktkommentar. EU-Datenschutz-Grundverordnung (EU-DSGVO) - Neues Bundesdatenschutzgesetz (BDSG) - Weitere datenschutzrechtliche Vorschriften. Bund-Verlag GmbH, Nördlingen, 2020.

[Gi21] Gierschmann, Sibylle: Gestaltungsmöglichkeiten durch systematisches und risikobasiertes Vorgehen – Was ist schon anonym? Zeitschrift für Datenschutz, 9:482–486, 2021.

[GLF18] Geminn, Christian; Laubach, Anne; Fujiwara, Shizuo: Schutz anonymisierter Daten im japanischen Datenschutzrecht. Zeitschrift für Datenschutz, S. 413–419, 2018.

[JE18] Janmey, V.; Elkin, P. L.: Re-Identification Risk in HIPAA De-Identified Datasets: The MVA Attack. AMIA Annu Symp Proc, 2018:1329–1337, 2018.

[KB20] Kühling, Jürgen; Buchner, Benedikt: Datenschutz-Grundverordnung/ Bundesdatenschutzgesetz. C.H.Beck, München, 2020.

[KS20] Kühling, Jürgen; Schildbach, Roman: Die Reform der Datentransparenzvorschriften im SGB V. Neue Zeitschrift für Sozialrecht, 2:41–50, 2020.

[K0] Kühling, Jürgen: Gesundheitsdatenschutzrecht im Zeitalter von „Big Data“. Datenschutz Datensicherheit - DuD, 44(3):182–188, März 2020.

[K1] Kühling, Jürgen: Der datenschutzrechtliche Rahmen für Datentreuhänder. Datenschutz Datensicherheit - DuD, 45(12):783–788, Dezember 2021.

[La16] Ladeur, Karl-Heinz: „Big Data“ im Gesundheitsrecht – Ende der „Datensparsamkeit“? Datenschutz Datensicherheit - DuD, 40(6):360–364, Juni 2016.

[Le13] Lejeune, Mathias: Datenschutz in den Vereinigten Staaten von Amerika. CR, S. 755–760, 2013.

[LLV07] Li, Ninghui; Li, Tiancheng; Venkatasubramanian, Suresh: t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In: 2007 IEEE 23rd International Conference on Data Engineering. S. 106–115, 2007.

[Ma06] Machanavajjhala, A.; Gehrke, J.; Kifer, D.; Venkitasubramaniam, M.: L-diversity: privacy beyond k-anonymity. In: 22nd International Conference on Data Engineering (ICDE'06). S. 24–24, 2006.

[Me21] Meyer, Stephan: Landesrechtliche Legaldefinitionen der „Anonymisierung“ im Anwendungsbereich der DS-GVO. Zeitschrift für Datenschutz, 12:669–673, 2021.

[MH20] Martini, Mario; Hohmann, Matthias: Der gläserne Patient: Dystopie oder Zukunftsrealität? Neue Juristische Wochenzeitschrift, 49:3573–3578, 2020.

[NAC07] Nergiz, Mehmet Ercan; Atzori, Maurizio; Clifton, Chris: Hiding the presence of individuals from shared databases. In: Proceedings of the 2007 ACM SIGMOD international conference on Management of data - SIGMOD '07. ACM Press, New York, New York, USA, 2007.

[PP21] Paal, Boris P.; Pauly, Daniel A.: Datenschutz-Grundverordnung Bundesdatenschutzgesetz. Beck C. H., München, 2021.

[RG21] Roßnagel, Alexander; Geminn, Christian L.: Vertrauen in Anonymisierung. Zeitschrift für Datenschutz, 9:487–490, 2021.

[Ro18] Rolfs, Christian; Giesen, Richard; Kreikebohm, Ralf; Udsching, Peter: BeckOK Sozialrecht. Verlag C.H. Beck, München, 2018.

[Ro19] Roßnagel, Alexander: Datenschutz in der Forschung. Zeitschrift für Datenschutz, S. 157–159, 2019.

[Ro21] Roßnagel, Alexander: Grundrechtsschutz in der Datenwirtschaft. Zeitschrift für Rechtspolitik, 6:173–175, 2021.

[SgDB20] Spiecker gen. Döhmann, Indra; Bretthauer, Sebastian: , Schutzlos in Karlsruhe: Neue Maßstäbe im einstweiligen Rechtsschutz und im Datenschutz vor dem Bundesverfassungsgericht. https://verfassungsblog.de/schutzlos-in-karlsruhe/, 2020. zuletzt aufgerufen am 08.07.2022.

[SHD19] Simitis, Spiros; Hornung, Gerrit; Döhmann, Indra Spiecker: Datenschutzrecht - DSGVO mit BDSG. Nomos Verlagsgesellschaft, Baden-Baden, 2019.

[Sp17] Spiecker gen. Döhmann, Indra: Big und Smart Data: Zweckbindung zwecklos? Spektrum der Wissenschaft, 2017(1):56–62, 2017.

[Sp20] Specht-Riemenschneider, Louisa; Buchner, Benedikt; Heinze, Christian; Thomsen, Oliver: Festschrift für Jürgen Taeger - IT-Recht in Wissenschaft und Praxis. Fachmedien Recht und Wirtschaft, 2020.

[St17] StiftungDatenschutz: Big Data und E-Health. Erich Schmidt Verlag GmbH Company, Berlin, 2017.

[Sw00] Sweeney, Latanya: Simple demographics often identify people uniquely. Health (San Francisco), 671:1–34, 2000.

[Sw02] Sweeney, Latanya: K-ANONYMITY: A MODEL FOR PROTECTING PRIVACY. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems, 10(05):557–570, Oktober 2002.

[SW17] Schantz, Peter; Wolff, Heinrich Amadeus: Das neue Datenschutzrecht. C.H.Beck, München, 2017.

[We20] Weichert, Thilo: "Datentransparenz" und Datenschutz. Medizinrecht, 38(7):539–546, Juli 2020.

[Zu21] Zuo, Z.; Watson, M.; Budgen, D.; Hall, R.; Kennelly, C.; Al Moubayed, N.: Data Anonymization for Pervasive Health Care: Systematic Literature Mapping Study. JMIR Med Inform, 9(10):e29871, Oct 2021.

Bereitstellung von Mobilitätsdaten – Ein Problem des Datenschutzes?

Betrachtung der Datenbereitstellungspflicht des Personenbeförderungsgesetzes aus der Perspektive von Open Data und dem Schutz personenbezogener Daten

Marc Widemann[1]

Abstract: An die Nutzung von Mobilitätsdaten sind vielfältige Hoffnungen zur Erreichung der Verkehrswende und für einen Beitrag gegen den Klimawandel geknüpft. Doch obwohl eine Vielzahl von Daten und Informationen ohnehin öffentlich bekannt gemacht werden muss, ist der Zugang zu Mobilitätsdaten an vielen Stellen noch herausfordernd. Neue Verkehrsformen und weitere Möglichkeiten zur Erhebung und Bereitstellung von Mobilitätsdaten erfordern in zunehmendem Maße, auch den Schutz personenbezogener Daten im Blick zu behalten. Wie ein Umgang mit Mobilitätsdaten aussehen kann, der gleichermaßen die Nutzbarkeit der Daten und den Schutz personenbezogener Daten sicherstellt, soll im folgenden Beitrag in Abgrenzung zu den Regelungen im Personenbeförderungsgesetz dargestellt werden.

Keywords: Mobilitätsdaten, Datenschutz, Anonymisierung, Personenbezug

1 Einleitung

Mobilitätsdaten[2] wird ein hoher Stellenwert in der Effizienzsteigerung des (öffentlichen) Verkehrs und damit auch der Bekämpfung des Klimawandels zugesprochen. So soll die Nutzung des mit vielen Problemen assoziierten motorisierten Individualverkehrs dadurch reduziert werden, dass Reisenden multi- oder intermodal verknüpfte Mobilitätsangebote zur Verfügung stehen. Ebenso kann die leichtere Verfügbarkeit von Informationen über Mobilitätsangebote abseits des motorisierten Individualverkehrs die Nutzungsschwelle senken. Eine Bereitstellung solcher Daten als offene Daten bietet sich an. Damit wird nicht nur die umfangreiche und kreative Nutzung dieser Daten ermöglicht und ihr Potenzial ausgeschöpft. Es entfällt auch der regelmäßige Abschluss von Nutzungsverträgen und weiterer administrativer Aufwand. Gleichzeitig liegen im Bereich der Mobilität Daten, die offensichtlich keinen Bezug zu Personen aufweisen – wie etwa Busfahrpläne – und Daten, die Auskunft über sehr sensible persönliche Informationen Auskunft geben – wie etwa einer Person zuzuordnende Fahrten – nah beieinander. Auch der Koalitionsvertrag der

[1] Rechtsanwälte BBG und Partner mbB, Contrescarpe 75A, 28195 Bremen, widemann@bbgundpartner.de.
[2] Der Begriff der Mobilitätsdaten wird in Anlehnung an das Personenbeförderungsgesetz verwendet und bezeichnet Daten über Angebote des öffentlichen Personennahverkehrs.

doi:10.18420/inf2022_46

2021 gewählten Bundesregierung sieht die freie Zugänglichkeit von Verkehrsdaten im Rahmen eines Mobilitätsdatengesetzes vor.[3]

Eine genaue Differenzierung der jeweiligen Daten danach, ob sie ausschließlich sachbezogen sind oder nicht, ist im Umgang mit Mobilitätsdaten unerlässlich, um den Schutz von Personen und die sinnvolle Verwendung von Daten gleichermaßen zu gewährleisten. Diese Aufgabe ist nur im Wege einer interdisziplinären Betrachtung zufriedenstellend zu lösen. Die Kenntnis des bestehenden Rechtsrahmens muss durch eine technische Perspektive ergänzt werden, aus der sich etwa Möglichkeiten zum Schutz aber auch Missbrauch von Daten ergeben.

2021 wurde das Personenbeförderungsgesetz (PBefG) geändert und unter anderem die Pflicht zur Bereitstellung von Mobilitätsdaten aufgenommen. Im folgenden Beitrag soll aufgezeigt werden, in welchem Verhältnis die Mobilitätsdaten zum Schutz personenbezogener Daten stehen und welche Folgen die unzureichende Trennung von personenbezogenen zu anderen Daten hat. Eine solche Trennung hätte die Nutzung der Daten und den Schutz personenbezogener Daten verbessert. Abschließend soll an einem konkreten Beispiel verdeutlich werden, dass eine solche Betrachtung aus rein juristischer Perspektive nicht ausreichend ist. Vielmehr müssen die technischen Möglichkeiten und damit verbundene Risiken bekannt sein, um eine juristische Bewertung vornehmen zu können.

2 Mobilitätsdaten als Open Data

Das für die Änderung des PBefG verantwortliche BMVI und die damalige Bundesregierung bekundeten ihre Absicht, Mobilitätsdaten als Open Data verfügbar zu machen und gleichzeitig personenbezogene Daten zu schützen.[4] Dabei übernahm die Bundesregierung die Definition von offenen Daten aus Erwägungsgrund (16) Richtlinie (EU) 2019/1024 über offene Daten und die Verwendung von Informationen des öffentlichen Sektors[5]. Danach sind Daten offen, wenn sie *„in einem offenen Format“* vorliegen und *„von allen zu jedem Zweck frei verwendet, weiterverwendet und weitergegeben werden können“*. Diesem selbstgesetzten Anspruch konnte die Änderung des PBefG hingegen nicht gerecht werden.

3 Deutscher Rechtsrahmen für (offene) Mobilitätsdaten im PBefG

Um die Bereitstellung von Daten in hinreichender Qualität sicherzustellen, wurden verschiedene Rechtsinstrumente geschaffen. Dazu gehören europäische Rechtsakte,

[3] [Me21], S. 41.
[4] [Po17], S. 6; [Op21], S. 21.
[5] ABl. L 172 vom 26.6.2019, S. 56–83.

Gesetze des Bundes und zum Teil auch Landesgesetze. Die folgende Darstellung wird sich jedoch auf die Regelungen des PBefG beschränken.

Obwohl für die Bereitstellung von Mobilitätsdaten bereits ein europäischer Rechtsrahmen existiert, wurden im Rahmen der Novellierung des PBefG im Jahr 2021 auch Vorgaben zur Datenbereitstellung aufgenommen, §§ 3a – 3c PBefG. Neben den Änderungen im PBefG werden (technische) Einzelheiten zur Datenbereitstellung und zum Datenabruf in der begleitend erlassenen und laufend angepassten Mobilitätsdatenverordnung (MDV) geregelt. Der Anwendungsbereich des PBefG ist indes auf die Personenbeförderung auf der Straße beschränkt. Damit sind jene Mobilitätsformen abseits der Straße, insbesondere auf der Eisenbahnschiene, und ohne Beförderungsleistung, beispielsweise das Car-, Scooter- oder Bikesharing, nicht erfasst.

3.1 Die Datenbereitstellung – § 3a PBefG

Gemäß § 3a Abs. 1 PBefG sind zur Datenbereitstellung Unternehmer[6] und Vermittler im Sinne des PBefG erfasst. Der Begriff des Unternehmers ergibt sich aus § 2 Abs. 1 i. V. m. § 3 PBefG, der Begriff des Vermittlers ist in § 1 Abs. 3 PBefG definiert. Die Daten sollen gemäß § 3a PBefG über den Nationalen Zugangspunkt (NAP; nach dem englischen Begriff national access point), § 2 Nr. 11 IVSG, bereitgestellt werden. In Deutschland übernimmt diese Aufgabe der „Mobilitäts Daten Marktplatz“ (MDM), im Juli 2022 wurde die *Mobilithek* verfügbar gemacht, sie soll den MDM sukzessive ablösen und schließlich ersetzen.[7]

Die bereitzustellenden Daten sind einerseits nach Daten des Linien- und Gelegenheitsverkehrs andererseits nach statischen und dynamischen Daten getrennt, § 3a Abs. 1 PBefG. Die Bereitstellungspflicht tritt gestaffelt in Kraft, seit Januar 2022 müssen alle statischen Daten bereitgestellt werden, seit Juli 2022 müssen auch die dynamischen Daten bereitgestellt werden, Art. 7 Abs. 2 Nr. 3 Gesetz zur Modernisierung des Personenbeförderungsrechts vom 16. April 2021. Einer Anpassung der MDV, die die Pflicht zur Bereitstellung dynamischer Daten umfasst, hat der Bundesrat am 10.06.2022 zugestimmt.[8]

Der Umfang der bereitzustellenden Daten wird in § 3a Abs. 1 PBefG bestimmt. Für den Linienverkehr[9] handelt es sich um folgende Daten:

- Statische Daten

[6] Begriffe wie Unternehmer, Vermittler und Anbieter beziehen sich in der absolut überwiegenden Mehrzahl auf Gruppen juristischer Personen, sodass eine genderneutrale Schreibweise dieser Begriffe unterbleibt.
[7] [Mo22].
[8] BR-Drs. 189/22.
[9] Der Begriff des Linienverkehrs im PBefG umfasst eine Reihe von Verkehrsarten, die ihrerseits nicht alle von der Datenbereitstellungspflicht erfasst sind. Zum leichteren Verständnis wird die Betrachtung auf klassische Linienverkehre, wie den fahrplangebundenen Bus- oder Straßenbahnverkehr, beschränkt.

  - Nr. 1 a) – Daten zum Anbieter und Angebot, dazu gehören Name und Kontaktdaten, Fahrpläne, Tarifstruktur, Buchungs- und Bezahlmöglichkeiten, Barrierefreiheit und Umweltstandard der eingesetzten Fahrzeuge
  - Nr. 1 c) – Daten zur Infrastruktur, dazu gehören Bahnhöfe[10], Haltestellen und andere Zugangsknoten, inklusive Daten zur Barrierefreiheit und Infrastruktur an den Zugangsknoten, wie Verkaufsstellen, Rolltreppen und Aufzüge
- Dynamische Daten
  - Nr. 1 b) – Daten zum Verkehrsangebot, dazu gehören Ausfälle, Störungen, prognostizierte Ankunfts- und Abfahrtszeit und die Auslastung des jeweiligen Verkehrsmittels
  - Nr. 1 d) – Daten zur Infrastruktur, dies umfasst den aktuellen Betriebsstatus der in Nr. 1 c) genannten Infrastruktur.

Für den Gelegenheitsverkehr[11] sind folgende Daten bereitzustellen

- Statische Daten
  - Nr. 2 a) – Daten zum Verkehrsangebot, dazu gehören Name und Kontaktdaten des Anbieters, Bediengebiete und -zeiten, Preise sowie Buchungs- und Bezahlmöglichkeiten und Daten zur Barrierefreiheit und zum Umweltstandard der eingesetzten Fahrzeuge
- Dynamische Daten
  - Nr. 2 b) – Daten zur Verfügbarkeit von Fahrzeugen an Stationen und im Verkehr inklusive deren Auslastung in Echtzeit sowie Daten zu den tatsächlich abgerechneten Kosten.

In der MDV wird die Pflicht zur Datenbereitstellung weiter konkretisiert. So wird etwa im Anhang der MDV detailliert aufgezählt, welche Daten genau bereitgestellt werden müssen und welche Standards für die Bereitstellung verwendet werden sollen. Die technische Ausgestaltung kann der NAP im Übrigen gemäß § 8 Abs. 1 MDV selbst bestimmen. Details dazu finden sich etwa in der Technischen Schnittstellenbeschreibung des MDM.[12]

[10] Das PBefG regelt ausschließlich die Personenbeförderung auf der Straße, Bahnhöfe des Eisenbahnverkehrs sind demnach nicht erfasst, sondern solche des Bus-, O-Bus- oder Straßenbahnverkehrs.

[11] Der Gelegenheitsverkehr umfasst eine Reihe von Verkehrsarten, die ihrerseits nicht alle von der Datenbereitstellungspflicht erfasst sind. Die bekanntesten Verkehrsarten des Gelegenheitsverkehrs dürften der Taxen- und Mietwagenverkehr sein. Das PBefG definiert den Begriff des Mietwagenverkehrs in § 49 Abs. 4 Satz 1 PBefG anders als im allgemeinen Sprachgebrauch. Beim Mietwagenverkehr ist die Beförderung – insoweit ähnlich wie beim Taxenverkehr – zwingend Teil der Leistung.

[12] [Te21].

3.2 Der Zugang zu den Daten – § 3b PBefG

Der Zugang zu den Daten richtet sich nach § 3b PBefG. In § 3b Abs. 1 und 2 PBefG sind eine Reihe von Stellen aufgelistet, denen der NAP in unterschiedlichem Umfang und zu verschiedenen Zwecken die bereitgestellten Daten übermittelt. Mit dem Recht, die Daten vom NAP übermittelt zu bekommen, geht gemäß § 3b Abs. 3 – 6 PBefG jeweils das Recht einher, die Daten zu erheben, zu speichern und zu verwenden, soweit dies zur Erfüllung des jeweiligen Zwecks erforderlich ist. Bei der überwiegenden Mehrzahl der berechtigten Stellen handelt es sich um staatliche Stellen, die in unterschiedlichem Umfang Daten zur Erfüllung verschiedener behördlicher Aufgaben erhalten. Private werden als *Dritte* bezeichnet. Diese Dritten erhalten Zugang zu allen Daten, beschränkt wird der Zugang jedoch auf die Zwecke „zur Erbringung bedarfsgesteuerter Mobilitätsdienstleistungen oder multimodaler Reiseinformationsdienste“[13] ,§ 3b Abs. 1 Nr. 3 PBefG.

Weitere Regelungen zum Zugang zu den Daten finden sich darüber hinaus in der MDV. So sind Dritte nach § 6 MDV verpflichtet, sich unter Angabe der dort aufgezählten Informationen, wie zustellungsfähige Anschrift und Telefonnummer, beim NAP zu registrieren. Darüber hinaus ergeben sich aus der Technischen Schnittstellenbeschreibung des MDM[14] weitere Vorgaben, die Dritte zum Bezug der Daten erfüllen müssen. Wird das *Brokering*[15] des MDM genutzt, ist demnach ein Bezug der Daten nur möglich, wenn sich die jeweiligen Systeme zum Datenabruf per vorher festgelegten Zertifikat am NAP identifizieren. Eine Datenbereitstellung ist auch dadurch möglich, dass einzelne Dateien durch die Datengeber manuell beim MDM hochgeladen werden und der MDM die Datei als Download zur Verfügung stellt. Dieser Link kann – sofern er öffentlich einsehbar sein soll – durch eine Suche beim MDM für den Datensatz des jeweiligen Anbieters gefunden werden.

Ebenso besteht die Möglichkeit, Dateien über eigene Systeme anzubieten und beim MDM lediglich die entsprechenden Metadaten inkl. URL zur Datenquelle zu hinterlegen. Ein direkter Zugriff auf die Dateien über die URL ist hierbei offenbar nicht notwendig. Da § 3b PBefG lediglich die Übertragung von Daten durch den NAP betrifft, hängt in diesen Fällen der Bezug davon ab, wie er beim jeweiligen Anbieter der Daten technisch ausgestaltet ist.

3.3 Offenheit der Mobilitätsdaten

Die damalige Bundesregierung hat mehrfach betont, Mobilitätsdaten als Open Data verfügbar machen zu wollen. Diesen Anspruch können die Regelungen im PBefG und der MDV nicht erfüllen. Abseits von öffentlichen Stellen haben nur Anbieter von bedarfsgesteuerten Mobilitätsangeboten und multimodalen Reiseinformationsdiensten einen Anspruch auf Datenübermittlung durch den NAP. Darüber hinaus ist die

[13] Beide Begriffe werden im PBefG nicht definiert, ihre genaue Bedeutung muss an dieser Stelle offenbleiben.
[14] [Te21].
[15] Hierbei werden die Daten auf die Systeme des MDM und von dort an die jeweiligen Empfänger übertragen.

Übermittlung der Daten im Fall des Brokering nur mittels einer technisch nicht trivialen Authentifizierung durch ein Zertifikat möglich.

Auch hinsichtlich der Verwendung der Daten zeigt sich das PBefG restriktiv. Gemäß § 3b Abs. 1 Nr. 3, Abs. 4 PBefG dürfen Private die Daten lediglich in einem solchen Umfang verwenden, soweit dies zur Erbringung von bedarfsgesteuerten Mobilitätsdienstleistungen und multimodalen Reiseinformationsdiensten erforderlich ist. Eine Verwendung der Daten zu anderen Zwecken ist untersagt und hat einen zwingenden Ausschluss vom Datenbezug durch den NAP zur Folge, § 5 Abs. 3 MDV. Ein Ermessen steht dem NAP hierbei nicht zu, sodass wohl selbst ein nützlicher Einsatz der Daten einen Ausschluss vom Datenbezug zur Folge haben dürfte.

4 Offene Mobilitätsdaten – ein Problem des Datenschutzes?

Als Grund für diese Einschränkungen wurde der Schutz personenbezogener Daten angeführt.[16] Ein solches Vorgehen wäre im Grundsatz absolut nachvollziehbar. Personenbezogene Daten können und dürfen von der Forderung nach Open Data nicht umfasst sein, da das berechtigte Schutzinteresse der Betroffenen das Interesse der Allgemeinheit an der Verwendung der Daten in aller Regel überwiegt. Der Chaos Computer Club hat dies bereits in den 1980er Jahren prägnant in dem Satz „Öffentliche Daten nützen, private Daten schützen“ auf den Punkt gebracht.[17]

Gleichzeitig drängt sich hinsichtlich der Daten über Infrastruktur oder Fahrpläne des öffentlichen Personennahverkehrs die Frage auf, wie eine Einschränkung ihrer Verwendung mit Datenschutzerwägungen gerechtfertigt werden kann.

4.1 Verhindert der Datenschutz offene Reise- und Verkehrsdaten?

Die Gesetzesbegründung gibt keinen Aufschluss darüber, welchen Personenbezug etwa ein Fahrplan aufweisen kann. In dieser wird zu § 3b PBefG pauschal darauf verwiesen, dass der NAP als „datenschutzrechtlich verantwortliche Stelle“[18] die Rechtmäßigkeit der Übermittlung personenbezogener Daten zu gewährleisten habe. Dafür sei es notwendig, dass der NAP Kenntnis über die Zwecke der Übermittlung habe. Vor diesem Hintergrund verwundert es nicht, dass § 3b PBefG hinsichtlich des Vokabulars – *erheben*, *verwenden* – und der Systematik an die Verordnung (EU) 2016/679[19] (DSGVO) erinnert. So lehnt sich offensichtlich auch die Festlegung eines numerus clausus von Zwecken, zu denen die

[16] BT-Drs. 19/26175, S. 32, 40.
[17] [Ch22].
[18] BT-Drs. 19/26175, S. 40.
[19] Verordnung (EU) 2016/679 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten, zum freien Datenverkehr und zur Aufhebung der Richtlinie 95/46/EG (Datenschutz-Grundverordnung), ABl. L 119 vom 4.5.2016, S. 1.

Daten übermittelt werden dürfen, an die Zweckbindung nach Art. 5 Abs. 1 lit. b) DSGVO an.

Es lohnt sich daher einen Blick auf die Natur der Mobilitätsdaten zu werfen. Das Regime der DSGVO etabliert wegen des engen Bezugs zu den Persönlichkeitsrechten der betroffenen Personen bewusst strenge Vorgaben, wie schon in Erwägungsgründen [1-4] DSGVO klargestellt wird. Gemäß Art. 2 Abs. 1 DSGVO ist diese jedoch ausschließlich auf personenbezogene Daten i. S. d. Art. 4 Nr. 1 DSGVO anwendbar. Es zeigt sich, dass auf eine Differenzierung der Daten nach personenbezogenen Daten und nach nichtpersonenbezogenen Daten, die nach Möglichkeit öffentlich zugänglich und frei verwendbar sind, verzichtet wurde. Die Folge ist, dass der gewünschten und sinnvollen Verwendung von Daten unnötig enge Grenzen gesetzt werden.

4.2 Personenbezogene Daten und Sachdaten

Der Abgrenzung von personenbezogenen Daten zu Sachdaten kommt hier also eine wichtige Rolle zu. Denn ohne Personenbezug ist der Schutzbereich der DSGVO bereits nicht eröffnet. Gleichzeitig führt das Vorliegen eines Personenbezugs nicht in jedem Fall dazu, dass die Bereitstellung entsprechender Daten verboten wäre. Vielmehr sind in Art. 6 UAbs. 1 lit. a) – f) DSGVO Bedingungen benannt, bei deren Erfüllung eine Verarbeitung rechtmäßig ist.

Gemäß Art. 4 Nr. 1 DSGVO handelt es sich bei personenbezogenen Daten um Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen. Davon, dass es sich um Informationen handelt, wird an dieser Stelle auch wegen der weiten Auslegung dieses Begriffs ausgegangen.[20] Die Merkmale des Bezugs zu einer natürlichen Person und die Frage der Identifizierbarkeit einer Person sind zudem getrennt zu betrachten (zu letzterem s. 5.2).

Entscheidend ist also der Bezug zu einer natürlichen Person. Denn im Umkehrschluss zu Art. 4 Nr. 1 DSGVO sind solche Daten – mangels Personenbezug – nicht geschützt, die sich ausschließlich auf Sachen beziehen (sog. Sachdaten).[21] Zu der Frage, wie Sachdaten von personenbezogenen Daten abzugrenzen sind, hat die Artikel-29-Datenschutzgruppe drei Kriterien ausgearbeitet.[22] Demnach sollen Daten dann als personenbezogen gelten, wenn ein Inhaltselement, ein Zweckelement oder ein Ergebniselement vorhanden sind. Für die Annahme eines Personenbezugs brauchen nicht alle drei Elemente vorzuliegen. Die Artikel-29-Datenschutzgruppe misst allerdings dem Inhaltselement eine besondere Rolle zu. Bei dessen Vorliegen käme es auf die anderen beiden Elemente nicht an, um

[20] Anstatt Vieler: Klar/Kühling in [KB20], Art. 4 Nr. 1, Rn. 9.
[21] Anstatt Vieler: Ernst in [PP21], Art. 4 DS-GVO, Rn. 6.
[22] [Ar07], S. 10ff.

einen Personenbezug anzunehmen.[23] Auch der EuGH bedient sich dieser drei Elemente, um festzustellen, ob Daten personenbezogen sind.[24]

Ein Inhaltselement gilt dann als vorhanden, wenn es sich bei den Informationen unter Berücksichtigung aller Begleitumstände um Informationen über eine Person handelt, also Informationen über eine Person gegeben werden. Dies gilt unabhängig vom Zweck einer Verarbeitung oder den Auswirkungen dieser Information auf die betroffene Person.

Werden Daten unter Berücksichtigung aller Begleitumstände mit dem Zweck verwendet, oder können sie mit dem Zweck verwendet werden, *„eine Person zu beurteilen, in einer bestimmten Weise zu behandeln oder ihre Stellung oder ihr Verhalten zu beeinflussen“*[25], soll ein Zweckelement vorliegen.

Beim Ergebniselement wird – wie der Name andeutet – darauf abgestellt, dass sich die Daten unter Berücksichtigung aller Begleitumstände schlussendlich auf die Rechte und Interessen einer bestimmten Person auswirken könnten. Diese Auswirkung muss nicht nachhaltig sein, ebenso lässt die Formulierung im Konjunktiv darauf schließen, dass bereits die Möglichkeit einer Auswirkung ausreichend sein dürfte.

4.3 Die verschiedenen Kategorien von Mobilitätsdaten

Auf dieser Grundlage lässt sich bei einigen Mobilitätsdaten leicht feststellen, dass kein Personenbezug gegeben ist. Dies betrifft vollständig die **statischen Daten zur Infrastruktur im Linienverkehr**[26]. Ebenso ist nicht erkennbar, wie sich ein Personenbezug aus **dynamischen Daten zum Betriebsstatus der Infrastruktur** im Linienverkehr[27] ergeben könnte.

Hinsichtlich der **übrigen statischen Daten im Linienverkehr**[28] ist zunächst danach zu unterscheiden, ob der Anbieter seinerseits vom Anwendungsbereich der DSGVO erfasst ist.[29] Denn nach Erwägungsgrund (14) Satz 2 DSGVO sollen personenbezogene Daten juristischer Personen nicht durch die DSGVO geschützt werden. Anders ist der Fall zu betrachten, wenn die Daten mit Bezug zu einer juristischen Person gleichzeitig einen Bezug zu einer natürlichen Person aufweisen.[30] Dies wird häufig dann der Fall sein, wenn der Name einer juristischen Person zugleich der Name einer natürlichen Person ist, wie es wohl hauptsächlich bei kleineren Unternehmen der Fall ist. Dem steht grundsätzlich nicht

[23] [Ar07], S. 13 f.
[24] [Eu17] Rn. 35 ff.
[25] [Ar07], S. 11 f.
[26] § 3a Abs. 1 Nr. 1 c) PBefG.
[27] § 3a Abs. 1 Nr. 1 d) PBefG.
[28] § 3a Abs. 1 Nr. 1 a) PBefG.
[29] Dies wäre grundsätzlich auch für die Daten zur Infrastruktur denkbar. Es wird aber davon ausgegangen, dass der Begriff von Infrastruktur von größeren juristischen Personen übernommen wird. Die Aufnahme der Infrastrukturdaten würde die folgende Bewertung zudem unberührt lassen.
[30] Anstatt Vieler: Klar/Kühling in [KB20] Art. 4 Nr. 1 Rn. 4 mit Verweis auf [Eu10], Rn. 52 ff.; Ebenso [EC22].

entgegen, dass sich die Daten vorliegend ausschließlich auf berufliche Tätigkeiten beziehen. Vom Begriff des Privatlebens, das nach Art. 7 Charta der Grundrechte der Europäischen Union[31] (GRCh) geschützt ist und einen engen Bezug zum Schutz personenbezogener Daten nach Art. 8 GRCh aufweist, welcher seinerseits durch die DSGVO konkretisiert wurde, können berufliche Tätigkeiten nicht ausgenommen werden.[32] Weisen die Angaben zum Anbieter der Verkehrsleistung keinen Bezug zu einer natürlichen Person auf, ist nicht ersichtlich, wie sich aus den übrigen statischen Daten des Linienverkehrs ein Personenbezug ergeben könnte. Ist ein solcher vermittelter Bezug zu einer natürlichen Person jedoch gegeben, würden demnach mindestens die Daten zum Namen des Anbieters von der DSGVO geschützt sein. Die Daten darüber, welche Busse eingesetzt oder welche Linien bedient werden, geben darüber hinaus Auskunft über die Auftragslage des jeweiligen Beförderungsunternehmens und somit über die – bspw. per Name – mit dem Unternehmen verknüpfte Person.

Es sind allerdings Zweifel angebracht, ob dieses Ergebnis im konkreten Fall sachgerecht ist. Bei den vorliegend umfassten Daten handelt es sich um Daten, die ihrerseits öffentlich und auch wesentlich im Interesse des jeweiligen Unternehmens bekannt sind. Im Urteil des EuGH[33] wurden die Zuwendungssummen erst aufgrund der angegriffenen Regelungen öffentlich und betrafen damit Daten, an deren Schutz ein grundsätzliches Interesse erkennbar ist.

Eine Ausweitung des Schutzes der DSGVO auf alle Daten, die in der dargestellten Weise einen vermittelten Bezug zu einer Person aufweisen, würde dazu führen, dass der – im Übrigen datenschutzrechtlich völlig unproblematische – Umgang mit Daten über Fahrzeuge oder Fahrpläne zur Einhaltung aller Vorgaben der DSGVO verpflichtete. Während die Verarbeitung selbst wohl regelmäßig gemäß Art. 6 UAbs. 1 lit. f) DSGVO gerechtfertigt wäre, würden Informations- und Schutzpflichten zu einer völlig unangemessenen Belastung führen. Diese Belastung hinge letztlich vom Zufall ab, nämlich davon, wie ein Verkehrsunternehmer sein Unternehmen benannt hat.[34]

Sachgerecht scheint es daher, Erwägungsgrund (14) DSGVO soweit Geltung zu verschaffen, dass ein „Durchschlagen“ der Informationen über juristische Personen auf natürliche Personen nur dann angenommen werden kann, wenn sich daraus auch ein tatsächlicher und legitimer Schutz der Privatheit einer natürlichen Person im Sinne der Art. 7 und 8 GRCh ergibt. Hiervon ist hinsichtlich der statischen Mobilitätsdaten im Linienverkehr nicht auszugehen. Die bereitzustellenden Daten sind ohnehin – auch im

[31] ABl. C 326 vom 26.10.2012, S. 391–407.
[32] [Eu10], Rn. 59.
[33] [Eu10].
[34] Dies lässt sich im vorliegenden Fall nicht über eine Anonymisierung (bspw. durch Entfernung des Namens des Unternehmers) der Daten lösen. Einerseits liegen die Daten am Nationalen Zugangspunkt zugänglich vor, sodass ein Personenbezug so leicht wiederherzustellen sein würde, dass eine Anonymisierung nicht möglich ist. Daneben besteht die Verpflichtung, die Daten unter Zuordnung zum jeweiligen Unternehmen zu verwenden, § 7 Nr. 1 MDV.

Interesse des Unternehmers – öffentlich zugänglich, eine irgendwie ausgeprägte Sensibilität ist bei ihnen ebenfalls nicht zu erkennen.

Hinsichtlich der Bereitstellung der **übrigen dynamischen** Daten ist zu differenzieren. Betrachtet man § 3a Abs. 1 Nr. 1 b) PBefG isoliert, ergibt sich das gleiche Bild, wie hinsichtlich der statischen Daten. Also, dass sich ein Personenbezug höchstens vermittelt über den Namen des Unternehmers ergeben kann. Allerdings ergibt sich aus dem Anhang der MDV, dass Daten zur voraussichtlichen Abfahrts- und Ankunftszeit dem Standard SIRI-VM entsprechend bereitgestellt werden sollen. Auch wenn die Bereitstellung von Positionsdaten in der MDV nicht vorgesehen ist, wird der Standard SIRI-VM für die Bereitstellung von Positionsdaten verwendet.[35] Eine überobligatorische Bereitstellung von Positionsdaten der Fahrzeuge ist demnach möglich. Hier ist zunächst zu unterscheiden. Für den jeweiligen Unternehmer handelt es sich unstreitig um personenbezogene Daten, denn anhand von Schichtplänen und ähnlichem ist dieser jederzeit und ohne relevanten Aufwand in der Lage, die Position der Fahrzeuge dem jeweiligen Fahrpersonal zuzuordnen. Die Schichtpläne und Identifizierungsmöglichkeit des Fahrpersonals sind allerdings nicht von der Datenbereitstellungspflicht erfasst. Auch im Übrigen ist davon auszugehen, dass diese Daten nicht öffentlich verfügbar sind. Sodass der Personenbezug durch Anonymisierung entfallen ist. Hierfür ist es nicht notwendig, dass niemand in der Lage wäre, einen Personenbezug herzustellen. Aus EG (26) DSGVO ergibt sich eine Bewertung anhand der jeweiligen Möglichkeiten des Verantwortlichen oder Dritter.[36] Grundsätzlich sind legale Möglichkeiten, wie eine zivilgerichtliche Klage, denkbar, um die Fahrzeuge dem jeweiligen Fahrpersonal zuzuordnen. Der damit verbundene Aufwand und die Einzelfallbezogenheit dürften jedoch so hoch sein, dass nicht davon ausgegangen werden kann, dass diese „wahrscheinlich zur Identifizierung“ genutzt werden.

Daneben ist ein Personenbezug gemäß den oben aufgestellten Kriterien wenigstens denkbar, im Ergebnis aber abzulehnen. Während die Daten nicht inhaltlich mit einer Person verknüpft sind, erscheint die Annahme nicht abwegig, dass die Daten zur Bewertung der Fahrer*innen – durch Andere als ihre Arbeitgeber*innen – herangezogen werden können oder ihr Verhalten zumindest mittelbar beeinflusst wird. Etwa, weil durch die Bereitstellung präziser Standortdaten festgestellt werden kann, ob die jeweilige Fahrer*in Geschwindigkeitsbeschränkungen beachtet.

Allerdings muss bei der Abgrenzung von Sach- zu personenbezogenen Daten auch die tatsächliche und relevante Erhöhung des Schutzes der Privatheit in die Überlegung miteinbezogen werden. So dürften sich Fahrer*innen öffentlicher Verkehrsmittel ohnehin permanent unter der Beobachtung ihrer Fahrgäste befinden und der Einfluss einer digitalen „Überwachung“ auf ihr Verhalten erwartbar gering bis verschwindend sein. Die Daten

[35] Dies lässt sich dem Abstract/Scope der entsprechenden Europäischen Norm EN 15531-3:2015 entnehmen, einzusehen u.a. unter https://standards.cencenelec.eu/dyn/www/f?p=CEN:110:0::::FSP_PROJECT:38718&cs=17E6D612292F4AA82B479CB55A3320E74, zuletzt abgerufen am 03.05.2022.

[36] Noch vor der DSGVO, aber mit gleichem Ergebnis: [Eu16].

betreffen zudem ausschließlich die Bewegung des Fahrzeugs, das sich ohnehin auf einem festen Linienweg bewegt.

Hinsichtlich der **statischen Daten im Gelegenheitsverkehr**[37] ergibt sich ein ähnliches Bild, wie bei den statischen Daten im Linienverkehr. Gleichwohl sprechen hier mehr Umstände für die Annahme eines Durchschlagens der unternehmensbezogenen Daten auf natürliche Personen. Dies ergibt sich daraus, dass die Unternehmer nicht selten deutlich kleinere Unternehmen führen und teilweise etwa zur geschäftlichen Kommunikation augenscheinlich auch eine private E-Mail-Adresse verwenden.[38]

Daraus kann sich ein Personenbezug auch bei **dynamischen Daten des Gelegenheitsverkehrs**[39] ergeben. Ebenso lassen sich die Überlegungen zu den Fahrer*innen im Linienverkehr auf die Fahrer*innen im Gelegenheitsverkehr übertragen. Hier kommt allerdings hinzu, dass der Gelegenheitsverkehr nicht auf vorgegebenen Linien fährt. So können mehr Situationen entstehen, in denen die Daten Kontroll- oder Einflussmöglichkeiten entwickeln oder eine Identifizierung des Fahrpersonal möglich wird, etwa bei einer Rückfahrt, nachdem ein Fahrgast ausgestiegen ist. Da nur Daten zu verfügbaren Fahrzeugen bereitzustellen sind, werden die tatsächlichen Fahrtwege der Fahrgäste im klassischen Taxen- und Mietwagenverkehr wohl nicht erfasst sein (zu möglichen Eingriffen in die Rechte der Fahrgäste s.u. 5).

4.4 Zwischenergebnis

Hier zeigt sich die Notwendigkeit genauer Differenzierung verschiedener Datenarten. So weist die absolute Mehrzahl der statischen Daten des Linienverkehrs im Grundsatz keinerlei Personenbezug auf, sodass einer freien Verwendung zumindest keine Gründe des Datenschutzes entgegenstehen. Ein vermittelter Personenbezug über den Namen des Unternehmers könnte sich in einigen Fällen ergeben. Gleichwohl spricht Vieles auch hier gegen die Annahme eines Personenbezugs.

Selbst wenn man einen solchen vermittelten Personenbezug annähme, widerspricht der Datenschutz der Bereitstellung von Mobilitätsdaten nicht grundsätzlich. Denn es ergäbe sich der Personenbezug allein durch die konkrete Ausgestaltung (Bereitstellung der Daten in Bezug zu Unternehmern, keine Nutzung der Daten ohne Bezug zu den Unternehmern). Hätte der Gesetzgeber die Datenbereitstellungspflicht anders ausgestaltet, etwa indem die Daten durch die Verbünde, nicht die Unternehmen, bereitgestellt werden müssen, hätte ein Personenbezug sicher ausgeschlossen werden können.

Im Gelegenheitsverkehr spricht mehr dafür, dass es sich in einer gewissen Zahl von Fällen um personenbezogene Daten handelt. Auch hier wäre es dem Gesetzgeber aber grundsätzlich möglich gewesen, durch Rechtsgestaltung den Personenbezug der eigentlich

[37] § 3a Abs. 1 Nr. 2 a) PBefG.
[38] Dies lässt sich aus den Daten schließen, die am „Mobilitäts Daten Marktplatz“ bereitgestellt werden.
[39] § 3a Abs. 1 Nr. 2 b) PBefG.

sachbezogenen Daten zu minimieren. Etwa durch eine getrennte Bereitstellung von ausschließlichen Sachdaten und möglicherweise personenbezogenen Daten.

Durch die fehlende Feststellung und Trennung von personenbezogenen und Sachdaten ergibt sich so eine – aus Datenschutzperspektive – unnötige Verwendungsbeschränkung. Aber auch für diejenigen, die schließlich mit den Daten arbeiten ergeben sich Rechtsunsicherheiten bezüglich der Verwendung der Daten. Dies betrifft etwa das nötige Schutzniveau und mögliche Auskunfts- und Informationsrechte.

5 Mögliche Auswirkung auf die Daten der Fahrgäste

Während die Verwendung von Sachdaten somit unter Verweis auf den Datenschutz erschwert wird, können sich zukünftig Situationen ergeben, in denen Daten der Fahrgäste nicht hinreichend geschützt werden.

Dies soll nachfolgend am Beispiel der Poolingverkehre verdeutlicht werden. Als Pooling wird das Bündeln verschiedener Fahrtwünsche der Fahrgäste entlang einer ähnlichen Strecke bezeichnet. Der Poolinganbieter sammelt die Start- und Zielwünsche der Fahrgäste und disponiert entsprechend Fahrzeuge, die die jeweiligen Fahrtwünsche möglichst effizient bedienen.

Die Poolingverkehre sind im PBefG als Linienbedarfsverkehr[40] (§ 44 PBefG), als gebündelter Bedarfsverkehr (§ 50 PBefG) oder als „ÖPNV-Taxi“ möglich. Beim ÖPNV-Taxi handelt es sich um eine besondere Form des Taxenverkehrs (§ 47 PBefG), bei dem einzelne Sitzplätze statt des gesamten Fahrzeugs vergeben werden. Der gebündelte Bedarfsverkehr und das ÖPNV-Taxi sind beide Teil des Gelegenheitsverkehres und sich hinsichtlich der sogleich aufgeworfenen Problemstellung so ähnlich, dass im Folgenden keine Unterscheidung zwischen beiden Verkehrsformen vorgenommen wird. Der Linienbedarfsverkehr unterliegt als Linienverkehr anderen Vorgaben zur Datenbereitstellungspflicht, daher wird sich die folgende Darstellung auf den Poolingverkehr des Gelegenheitsverkehrs beschränken.

5.1 Rechtliche Ausgangslage

Wie dargestellt, werden Unternehmer und Vermittler gemäß § 3a Abs. 1 Nr. 2 b) PBefG verpflichtet, dynamische Daten des Gelegenheitsverkehrs über den Nationalen Zugangspunkt bereitzustellen. Dazu gehören „Daten zur Verfügbarkeit von Fahrzeugen an Stationen und im Verkehr inklusive deren Auslastung in Echtzeit“, gemäß Anhang der MDV handelt es sich dabei im Einzelnen um:

[40] Die Möglichkeit einer Haustür-zu-Haustürbedienung ist für den Linienbedarfsverkehr umstritten.

- Geokoordinaten, Fahrzeugtyp und Barrierefreiheit von für eine Fahrt verfügbaren Fahrzeugen unter Verwendung der Ordnungsnummer in Echtzeit und
- den Auslastungsgrad, angegeben durch freie Sitzplätze und freie barrierefreie Sitzplätze im Fahrzeug.

Unter „Geokoordinaten“ ist die per GPS ermittelte genaue Position des Fahrzeugs zu verstehen.

Im klassischen Taxen- oder Mietwagenverkehr werden zu einem Fahrzeug, in dem ein Fahrgast sitzt, diese Daten nicht zu übermitteln sein, da diese in der Regel[41] „im Ganzen“ gemietet werden und das Fahrzeug damit nicht mehr als verfügbar gilt. Dagegen werden über Fahrzeuge des Poolingverkehrs nahezu dauerhaft Standortdaten und Daten zur Auslastung bereitzustellen sein. Wie sich die Verfügbarkeit für eine Fahrt im Poolingverkehr im Einzelnen abgrenzen lässt, muss an dieser Stelle offenbleiben. Zur vereinfachten Darstellung wird von der – im Übrigen auch naheliegenden – Annahme ausgegangen, dass nur dann Daten zu einem Fahrzeug zu liefern sind, wenn es noch über freie Sitzplätze verfügt.

Diese Daten werden gemäß § 3b Abs. 1 Nr. 3 PBefG vom Nationalen Zugangspunkt an Private zur Erbringung bedarfsgesteuerter Mobilitätsdienstleistungen oder multimodaler Reiseinformationsdienste übermittelt. Gemäß § 3b Abs. 1 Nr. 3 2. Halbsatz PBefG darf der Nationale Zugangspunkt diese Daten nicht übermitteln, *„wenn er Kenntnis davon erlangt, dass auf Grund der besonderen Umstände des Einzelfalls die Gefahr besteht, dass mit diesen Daten Bewegungen oder Standorte individualisierbarer Personen nachvollzogen werden können und diese Personen nicht in die Übermittlung eingewilligt haben“*.

5.2 Personenbezug der Geokoordinaten der Fahrzeuge

Bei den Daten über die Geokoordinaten handelt es sich grundsätzlich, anders als bei den meisten anderen Daten, nahezu immer um personenbezogene Daten. Denn sie geben Auskunft über den Aufenthaltsort einer bestimmten Person zu einem bestimmten Zeitpunkt. Durch Zusammenfassung der einzelnen Punkte ergibt sich, von welchem zu welchem Ort eine Person gereist ist. Werden die Daten aus der Perspektive des Anbieters des Poolingverkehrs betrachtet, wird dies besonders deutlich. Der Anbieter kann einerseits über die Buchungen, andererseits aber auch durch den tatsächlichen Fahrtantritt der Fahrgäste wissen, welche Personen zu welchem Zeitpunkt von welchem zu welchem Ort gefahren sind.

[41] Im Verkehr mit Mietwagen ist ausschließlich das Mieten von Fahrzeugen „im ganzen“ erlaubt, § 49 Abs. 4 PBefG.

Die Namen und auch andere Angaben zur Unterscheidung der Fahrgäste sind von der Datenbereitstellungspflicht allerdings nicht umfasst. Daher stellt sich die Frage, ob die Daten anonymisiert sind und der Personenbezug dadurch entfällt, EG (26), Art. 4 Nr. 1 DSGVO. Daten dürften dann als anonymisiert gelten, wenn der Personenbezug nur noch mit einem unverhältnismäßigen Aufwand an Zeit, Kosten und Arbeitskraft wiederhergestellt werden kann.[42] Dabei sind gemäß EG (26) DSGVO alle Mittel zu berücksichtigen, deren Verwendung nach allgemeinem Ermessen wahrscheinlich genutzt werden. Die Artikel-29-Datenschutzgruppe ging dann von einer wirksamen Anonymisierung aus, wenn es nicht möglich ist, eine einzelne Person herauszustellen, Verbindungen zwischen zwei Einträgen eines Datenbestands (oder zwei verschiedenen Datenbeständen) herzustellen oder Informationen aus den Datenbeständen abzuleiten.[43] Eine Identifikation anhand des Namens einer Person ist dagegen nicht notwendig, ausreichend ist auch eine Identifikation anhand anderer Merkmale.[44]

Davon ausgehend sind die Daten, die im Gelegenheitsverkehr bereitzustellen sind und Poolingverkehre betreffen, nicht durchgängig anonymisiert, sondern weiterhin personenbezogen. Durch die kombinierte Auswertung von freien Plätzen und der Positionsangabe des Fahrzeugs lässt sich aus den Daten leicht entnehmen, wann und wo eine Person ein- und ausgestiegen ist. Im einfachsten Fall, wenn lediglich ein Fahrgast fährt, lässt sich damit unmittelbar die Reiseroute dieser Person feststellen. Außerhalb sehr dicht besiedelter Gebiete lässt sich – sofern der Start oder das Ziel an einem Wohnhaus liegen – so bereits eine begründete Vermutung darüber anstellen, welche Reise ein*e Bewohner*in des jeweiligen Gebäudes getätigt hat. Kommen weitere Umstände hinzu – wiederholte Fahrten, ungewöhnliche Fahrtzeiten, sehr dünne Besiedelung – wächst auch die Wahrscheinlichkeit, dass die jeweilige Person tatsächlich dort wohnt und nicht etwa zu Besuch war oder zufällig von oder nach dort eine Reise unternommen hat.

Auch bei Daten, bei denen mehrere Personen das Poolingfahrzeug genutzt haben, lassen sich Annahmen ableiten. Dies sei an folgendem Fall verdeutlicht:

Es nutzen zwei Personen das Fahrzeug, wobei eine Person bei einem Studierendenwohnheim, die andere Person in einer von Einfamilienhäusern geprägten Gegend zugestiegen ist, sodass sich zeitweise beide Personen gemeinsam im Fahrzeug befinden. Steigen diese Personen nun einerseits an einem traditionell bei Studierenden beliebten Lokal, andererseits an einer Fachklinik für Dialyse aus, liegt die Vermutung nahe, welche Person welche Reise getätigt hat. Es ist zu befürchten, dass sich mit einer ausreichend großen Sammlung von Daten und der geschickten Auswertung dieser Daten in einer Vielzahl von Fällen, Informationen über die Reisenden gewinnen lassen.[45] Werden die Geokoordinaten vom Ein- und Ausstieg mit (in der Regel öffentlich

[42] [Bu20], S. 4.
[43] [Ar14], S. 9.
[44] [Ar07], S. 16.
[45] Bereits 2014 konnte Vijay Pandurangan einen – wenngleich anders aufgebauten – 20GB großen Datensatz mit 173 Mio Taxifahrten auswerten und die Fahrer*innen deanonymisieren, https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1, zuletzt abgerufen am 03.02.2022.

zugänglichen) Informationen über besonders sensible Orte (fachärztliche Einrichtungen, Rechtsanwält*innen für Straf- oder Familienrecht, Bordelle, Treffen von Selbsthilfegruppen und so weiter) kombiniert, lassen sich begründete Annahmen über höchstpersönliche Lebensbereiche der Fahrgäste treffen.

Dies gilt umso mehr für die Nutzer*innen der barrierefreien Plätze. Diese sind von den anderen Fahrgästen bei einer Auswertung leicht zu unterscheiden. Die Tatsache, dass eine Person einen barrierefreien Platz nutzt, hebt sie bereits deutlich von anderen Nutzer*innen des Fahrzeugs und auch ein Stück weit von der Gesamtbevölkerung ab. Somit lassen sich über diese Personen besonders leicht begründete Annahmen treffen. Denn sie stechen aus dem übrigen Wechsel der Fahrgäste wegen der getrennten Angabe immer heraus. Gleichzeitig ist zu erwarten, dass die Zahl der Fahrgäste, die einen barrierefreien Platz in Anspruch nehmen deutlich geringer ist. Außerdem sind diese Personen selbst bei dichterer Besiedelung möglicherweise leicht zu identifizieren, da zu den unspezifischen Annahmen (wo die Person möglicherweise wohnt und wo sie möglicherweise hinfährt) eine sehr spezifische Angabe (die Nutzung eines barrierefreien Platzes) hinzukommt. Weiterhin dürfte es sich bei der Angabe auch um ein besonders geschütztes Gesundheitsdatum i. S. d. Art. 4 Nr. 15, Art. 9 Abs. 1 DSGVO handeln. Gelingt die wahrscheinliche Zuordnung von Fahrten zu einer Person, ergeben sich daraus nicht nur Kenntnisse über die Gewohnheiten dieser Person, sondern zwingend auch über ihre Gesundheit.

Ein solches Vorgehen dürfte auch nicht an technischen oder finanziellen Hürden scheitern. Die Speicherung von Datenmengen im Bereich mehrerer Terrabyte ist selbst für Private erschwinglich. Die zur Auswertung nötige Rechenleistung dürfte ebenfalls bereits bei privat genutzten Computern vorhanden sein oder könnte zumindest zu vertretbaren Preisen gemietet werden.

Ein wirksamer Schutz gegen diese Auswertung ergibt sich aus den Vorgaben der §§ 3a ff. PBefG und der MDV dagegen nicht, auch wenn sich Private, die die Daten abrufen wollen, am Nationalen Zugangspunkt registrieren müssen, § 6 MDV. Eine umfassende Prüfung wird spätestens dann nicht mehr möglich sein, wenn das Datenangebot – wie vorgesehen – häufig und innerhalb der Europäischen Union nachgefragt und genutzt wird. Eine Vielzahl von genauen Überprüfungen wird durch den Nationalen Zugangspunkt nicht leistbar sein. Eine Auswertung der Daten ist zwar gemäß § 3c Abs. 2 PBefG verboten und dürfte gegen die DSGVO verstoßen, das Entdeckungsrisiko ist allerdings so gering, dass nicht von einer hinreichenden Abschreckungswirkung ausgegangen werden kann.

Ebenso bleibt offen, welchen Schutz § 3b Abs. 1 Nr. 3 2. Halbsatz PBefG entfalten soll und kann. Dafür, dass eine Datenübertragung unterbleibt, müsste der Nationale Zugangspunkt Kenntnis von „besonderen Umständen des Einzelfalls“ erlangen. Wie so eine Kenntnis erlangt werden könnte, ist unklar, denn der Nationale Zugangspunkt arbeitet selbst nicht mit den bereitgestellten Daten. Zudem geht die Gefahr hier gerade nicht von einzelnen Daten(sätzen) aus, sondern von der kontinuierlichen Sammlung und Auswertung großer Datenmengen. Unklar ist daher ebenfalls, ab welchem Zeitpunkt eine

hinreichend konkrete Gefahr vorliegen soll.[46] Ginge man von einer nachträglichen Bewertung aus, wären die Daten bereits übermittelt und die Entstehung des Schadens könnte nicht mehr verhindert werden. Sollen zukünftige Gefahren verhindert werden, müsste die Pflicht zur Bereitstellung von Positionsdaten mindestens im Pooling unangewendet bleiben.

6 Ergebnis

Die angeführten Beispiele verdeutlichen die Relevanz eines bewussten Umgangs mit Daten. Die mangelnde Feststellung, welche Daten personenbezogen sind, führt so einerseits dazu, dass eine freie und damit möglichst sinnvolle Verwendung von Sachdaten beschränkt wird. Hinsichtlich möglicherweise personenbezogener Daten (Fahrer*innendaten und Unternehmer*innendaten) führt dieses Vorgehen zu Rechtsunsicherheiten und möglichen Schutzlücken. Denn die Verwender*innen der Daten stehen ihrerseits vor der Entscheidung, die Daten als personenbezogen zu behandeln, was den oben angerissenen Mehraufwand nach sich zieht. Oder die Daten als Sachdaten zu behandeln, was – sollte tatsächlich ein Personenbezug vorliegen – zu Eingriffen in die Rechte der Betroffenen führen kann.

Gleichzeitig zeigt sich in den Regelungen der Anspruch, die Daten „irgendwie“ verwendbar zu machen. Dies führt dazu, dass erhebliche Eingriffe in die Rechte der Fahrgäste zu befürchten sind. Es wurde offenbar pauschal versucht, die Verhinderung dieser Eingriffe von der juristischen in die technische Sphäre zu verschieben, indem die Übertragung in einigen Fällen untersagt wird. Dieser Versuch muss allerdings als gescheitert gelten.

Hier hätte eine vorherige Auseinandersetzung stattfinden müssen, ob den Daten ein Personenbezug innewohnt. Darauf aufbauend hätten die Regelungen einerseits Konstellationen vermeiden können, in denen ein Personenbezug anzunehmen oder zu befürchten ist. Andererseits hätten für die Daten je nach Vorliegen eines Personenbezugs unterschiedliche Regeln geschaffen werden müssen, die den Eigenheiten der jeweiligen Daten entsprechen. Vorgaben zum Umgang mit diesen Daten müssen dem notwendigen Schutz personenbezogener Daten Rechnung tragen. Sie sollten sich aber auch auf personenbezogene Daten beschränken, um die Verwendung übriger Daten nicht zu erschweren. Eine klare Zuordnung von Daten zu den Kategorien personenbezogen oder sachbezogen stellt den Schutz der Betroffenen und die rechtssichere Verwendung der Daten sicher.

Dabei braucht es die Betrachtung von Daten aus technischer und juristischer Perspektive, um feststellen zu können, bei welchen Daten es sich ausschließlich um Sachdaten handelt und bei welchen durch technische Maßnahmen ein Personenbezug herstellbar ist. Die

[46] Hinzu kommt, dass das Verbot der Speicherung und Auswertung rechtmäßig Handelnden die Feststellung erschwert, dass ein Missbrauchspotenzial vorliegt.

Frage, ob Daten anonymisiert sind, kann nur interdisziplinär und nach eingängiger Prüfung beantwortet werden. Für den Missbrauch scheinbar anonymisierter Daten ist es ausreichend, einen Weg zu finden, aus den Daten Informationen über Personen abzuleiten. Sind die Daten erst verfügbar, haben entsprechende Angreifer*innen nahezu unbegrenzt Zeit und unbegrenzt viele Versuche ihr Ziel zu erreichen. Um den – auch zukünftigen – Missbrauch solcher Daten zu verhindern, ist dagegen ein Ausschluss möglichst aller Wege notwendig, über die Informationen zu Personen aus Daten abgeleitet werden. Dies wird ohne umfassende Kenntnis von technischen Mitteln und Möglichkeiten kaum möglich sein. Daher muss im Entwicklungsprozess entsprechender Rechtssetzung technische Expertise eingeholt werden, um Möglichkeiten und Gefahren der jeweiligen Daten hinreichend bewerten zu können.

Quellenverzeichnis

[Ar07] Artikel-29-Datenschutzgruppe, Stellungnahme 4/2007 zum Begriff „personenbezogene Daten“, WP136, 20.06.2007.

[Ar14] Artikel-29-Datenschutzgruppe, Stellungnahme 05/2014 zu Anonymisierungstechniken, WP216, 10.04.2014.

[Bu20] Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit, Positionspapier zur Anonymisierung unter der DSGVO unter besonderer Berücksichtigung der TK-Branche, 29.06.2020, abrufbar unter https://www.bfdi.bund.de/SharedDocs/Downloads/DE/Konsultationsverfahren/1_Anonymisierung/Positionspapier-Anonymisierung.pdf?__blob=publicationFile&v=4

[Ch22] Chaos Computer Club – Hackerethik, https://www.ccc.de/de/hackerethik, Stand: 29.04.2022.

[EC22] Europäische Kommission, https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations/application-regulation/do-data-protection-rules-apply-data-about-company_de, Stand: 21.04.2022.

[Eu10] EuGH, Urt. v. 09.11.2010, C-92/09 & 93/09.

[Eu16] EuGH, Urt. v. 19.10.2016, C-582/14, Breyer.

[Eu17] EuGH, Urt. v. 20.12.2017, C-434/16, Nowak.

[KB20] Kühling, J.; Buchner, B.: Datenschutz-Grundverordnung, Bundesdatenschutzgesetz: DS-GVO / BDSG, 3. Auflage, C.H.Beck, 2020.

[Me21] Mehr Fortschritt Wagen – Bündnis für Freiheit, Gerechtigkeit und Nachhaltigkeit – Koalitionsvertrag 2021 – 2025 zwischen der Sozialdemokratischen Partei Deutschlands (SPD), Bündnis 90 / Die Grünen und den Freien Demokraten (FDP), Berlin 2021.

[Mo22] Bundesministerium für Digitales und Verkehr, https://bmdv.bund.de/SharedDocs/DE/Artikel/DG/mobilithek.html, Stand. 29.04.202.

[PP21] Paal, B.; Pauly D.: Datenschutz-Grundverordnung Bundesdatenschutzgesetz: DS-GVO BDSG, 3. Auflage, C.H.Beck, 2021.

[Po17] Positionspapier Open Data des BMVI, Version 1.0, 24. Mai 2017.

[Op21] Open-Data-Strategie der Bundesregierung, Juli 2021.

[Te21] Technische Schnittstellenbeschreibung – Version 3.0.1. – 24.08.2021, abrufbar unter https://www.mdm-portal.de/wp-content/uploads/2022/03/MDM-Technische-Schnittstellenbeschreibung-V3.0.1.pdf.

Zulässigkeit von Open Source-Ermittlungen zur Strafverfolgung im Darknet

Sandra Wittmer[1] und Florian Platzer[2]

Abstract: Die Möglichkeiten für Cyberkriminelle, Straftaten im Internet anonym zu begehen, haben mit der Entwicklung von Darknet-Technologien erheblich zugenommen. Zur Aufklärung dieser Straftaten kommt aus technischer Perspektive die Nutzung von Open Source Intelligence (OSINT) in Betracht. Diese Methode bezeichnet einen Vorgang, bei dem Daten aus öffentlich zugänglichen Quellen gesammelt werden, um durch deren Auswertung an ermittlungsrelevante Erkenntnisse zu gelangen. Der folgende Beitrag thematisiert die Nutzung von OSINT zur Strafverfolgung im Darknet und wirft die Frage auf, was nach dem aktuellen Stand der Technik möglich und was auf Grundlage der geltenden strafprozessualen Vorgaben rechtlich zulässig ist. Die Verfasser*innen kommen zu dem Ergebnis, dass ein legislatives „Update" der StPO unvermeidbar sein wird, wenn die rechtlichen Vorgaben mit den technischen Möglichkeiten Schritt halten sollen. Dieses Ergebnis wird damit begründet, dass die Nutzung von OSINT aufgrund ihrer grundrechtlichen Eingriffsintensität in vielen Fällen nicht mehr auf die Ermittlungsgeneralklausel der §§ 161, 163 StPO gestützt werden kann.

Keywords: Open Source Intelligence, Strafverfolung, Cybercrime, Darknet

1 Einleitung

Während der Begriff Darknet bei den meisten Menschen bedrohliche Assoziationen hervorruft, ist es für andere ein Ort, dem in Zeiten der Snowden-Enthüllungen rund um die umfangreichen Überwachungs- und Spionagetätigkeiten der NSA die Hoffnung auf ein „besseres" Internet zugrunde liegt. Unbestritten ist, dass die Möglichkeiten für Cyberkriminelle, Straftaten im Internet anonym zu begehen, mit der Entwicklung von Darknet-Technologien erheblich zugenommen haben. Als bekanntes Beispiel hierfür kann die Anonymisierungssoftware „Tor"[3] genannt werden, die es ihren Nutzer*innen ermöglicht, Internetdienste zu betreiben und zu benutzen, ohne dabei ihre IP-Adressen preisgeben zu müssen. Erstmals in den Fokus der öffentlichen Wahrnehmung gerückt sind Darknet-Technologien, nachdem im Juli 2016 die Hintergründe des Attentats vor dem

[1] Fraunhofer SIT/ATHENE, Rheinstraße 75, 64295 Darmstadt, sandra.wittmer@sit.fraunhofer.de
[2] Fraunhofer SIT/ATHENE, Rheinstraße 75, 64295 Darmstadt, florian.platzer@sit.fraunhofer.de, https://orcid.org/0000-0003-0851-7744.
[3] Vgl. Tor Project, https://www.torproject.org (letzter Zugriff 29.04.2022).

doi:10.18420/inf2022_47

Olympia-Einkaufszentrum in München ans Licht gekommen waren.[4] In den darauffolgenden Jahren wurde eine rege Diskussion über das Missbrauchspotential von Darknet-Technologien geführt, die hierzulande sogar in der Forderung gipfelte, deren Verwendung in freiheitlich-demokratischen Staaten gänzlich zu verbieten.[5] Obwohl es dazu aus technischen und auch (verfassungs-)rechtlichen Gründen nicht gekommen ist,[6] wurde auf strafprozessualer Ebene intensiv nach Möglichkeiten gesucht, die strafrechtliche Verfolgung von darknetspezifischer Kriminalität voranzutreiben. Besonders solche Ermittlungsansätze, die eine digitale Spurensuche in Form der automatisierten Sammlung und Auswertung großer Datenmengen umfassen (sogenannte Open Source-Ermittlungen), sind dabei in den Fokus der Ermittlungsbehörden geraten. Diese sind bei der Strafverfolgung im Darknet besonders aussichtsreich, da „klassische" Ermittlungsinstrumente dort infolge der Anonymisierung von IP-Adressen erfolglos bleiben. Um dennoch Anhaltspunkte für die Identifizierung von tatverdächtigen Personen zu gewinnen, werden daher alternative Ermittlungsansätze benötigt. Als aktuelles Praxisbeispiel für den Einsatz solcher alternativer Ermittlungsmethoden bei der Strafverfolgung im Darknet kann die von der niederländischen Forschungseinrichtung TNO entwickelte Software „Dark Web Monitor" genannt werden.[7] Seit Juli 2020 wird das Tool, das öffentlich zugängliche Informationen aus dem Darknet sammelt und im Anschluss zu Strafverfolgungszwecken analysiert, durch die Zentralstelle Cybercrime Bayern (ZCB) in Bamberg getestet. Laut dem Bayerischen Staatsminister der Justiz Georg Eisenreich hat sich die Software schon jetzt als „wertvoll erwiesen".[8] Angekündigt wurde zudem, dass von dieser Technologie Ermittler*innen in ganz Deutschland und darüber hinaus auch in Europa profitieren werden.[9] Vor dem Hintergrund dieser Entwicklungen

---

[4] Ein 18-jähriger Schüler hatte dort mit einer Waffe, die er über das deutschsprachige Darknet-Forum „Deutschland im DeepWeb" erworben hatte, neun Menschen erschossen und fünf weitere verletzt, vgl. *Grundner/Rauch*, OEZ-Attentat in München: Als der Hass neun Menschen tötete, BR 24 vom 22.07.2021, abrufbar unter https://www.br.de/nachrichten/bayern/oez-attentat-in-muenchen-als-der-hass-neun-menschen-toetete,SdniNXM (letzter Zugriff: 29.04.2022).

[5] Vgl. *Borchers*, Europäischer Polizeikongress: Weg mit dem Darknet, Heise Online (20.02.2019), abrufbar unter https://www.heise.de/newsticker/meldung/Europaeischer-Polizeikongress-Weg-mit-dem-Darknet-4313276.html (letzter Zugriff: 30.01.2022).

[6] Vgl. *Wittmer* et al., Deanonymisierung im Tor-Netzwerk – Technische Möglichkeiten und rechtliche Rahmenbedingungen, in: Friedewald/Kreutzer/Hansen (Hrsg.), Selbstbestimmung, Privatheit und Datenschutz. Gestaltungsoptionen für einen europäischen Weg, S. 154.

[7] Vgl. *Bayerisches Staatsministerium der Justiz*, Mehr Licht ins Darknet: Dark Web Monitor soll Strafverfolgungsbehörden bei Ermittlungen im Darknet verstärken, Pressemitteilung vom 27.07.2020, abrufbar unter https://www.justiz.bayern.de/presse-und-medien/pressemitteilungen/archiv/2020/69.php (letzter Zugriff: 30.01.2022).

[8] Vgl. *Bayerisches Staatsministerium der Justiz*, Mehr Licht ins Darknet: Dark Web Monitor soll Strafverfolgungsbehörden bei Ermittlungen im Darknet verstärken, Pressemitteilung vom 27.07.2020, abrufbar unter https://www.justiz.bayern.de/presse-und-medien/pressemitteilungen/archiv/2020/69.php (letzter Zugriff: 30.01.2022).

[9] Vgl. *Bayerisches Staatsministerium der Justiz*, Mehr Licht ins Darknet: Dark Web Monitor soll Strafverfolgungsbehörden bei Ermittlungen im Darknet verstärken, Pressemitteilung vom 27.07.2020, abrufbar unter https://www.justiz.bayern.de/presse-und-medien/pressemitteilungen/archiv/2020/69.php (letzter Zugriff: 30.01.2022). Neben dem von der TNO entwickelten „Dark Web Monitor" existieren inzwischen zudem auch zahlreiche kommerzielle Anbieter*innen wie „Hunchly" (eine Online-Demo des

sollen im nachfolgenden Beitrag Open Source-Ermittlungen zur Strafverfolgung im Darknet untersucht und die Frage aufgeworfen werden, was nach dem aktuellen Stand der Technik möglich und was auf Grundlage der geltenden strafprozessualen Vorgaben rechtlich zulässig ist. Dabei werden in erster Linie (deutsche) strafprozessuale Belange behandelt und unions- sowie datenschutzrechtliche Vorgaben thematisiert. Anschließend wird der Frage nachgegangen, anhand welcher Kriterien die Grenze zwischen strafprozessual gerechtfertigten Open Source-Ermittlungen einerseits und der unzulässigen Datensammlung und -Auswertung andererseits gezogen werden kann. Das übergeordnete Ziel des Beitrags soll es dabei sein, eine Diskussionsgrundlage für mögliche Vorgehensweisen bei der Strafverfolgung im Darknet zu schaffen, ohne jedoch die Legitimität und Legalität einer anonymen Nutzung des Internets grundsätzlich in Frage zu stellen.

2 Terminologie und technische Grundlagen

Der Begriff der Open Source Intelligence (OSINT) stammt ursprünglich aus der Welt der Nachrichtendienste und bezeichnet einen Vorgang, bei dem zum Zwecke der Informationsgewinnung Daten aus frei verfügbaren Quellen gesammelt werden, um durch deren Auswertung und Analyse an übergeordnete, ermittlungsrelevante Erkenntnisse zu gelangen.[10] Im folgenden Beitrag wird der Begriff „Open Source-Ermittlungen“ verwendet, um den Einsatz von OSINT zur Informationsgewinnung in strafrechtlichen Ermittlungsverfahren von Konstellationen abzugrenzen, in denen OSINT zu nachrichtendienstlichen Zwecken genutzt wird und demzufolge das BNDG und nicht die StPO Anwendung findet.

2.1 Begriff und Methoden der Open Source Intelligence (OSINT)

Kommen Open Source-Ermittlungen zu Strafverfolgungszwecken zum Einsatz, können sämtliche Informations- und Datenquellen genutzt werden, die sowohl offline als auch im Internet öffentlich einsehbar sind und für deren Beschaffung keine besondere Legitimation benötigt wird.[11] Als Beispiel hierfür können mediale Quellen (Zeitung, Radio, Fernsehen), webbasierte Communities (soziale Netzwerke, Videoplattformen, Wikis, Foren), staatliche Quellen (Regierungsberichte, Anhörungen, Pressekonferenzen), akademische Quellen (Konferenzen, akademische Paper) oder georäumliche Quellen (Karten,

Tools kann unter https://www.hunch.ly/osint-webinars/osint-webinar-dark-web-1252018 (letzter Zugriff: 30.011.2022) eingesehen werden) oder „Recorded Future“, die ebenfalls Ermittlungstools für das Darknet entwickelt haben.

[10] Zum Begriff der Open Source Intelligence siehe *Deusch/Eggendorfer*, in: Taeger/Pohle, ComputerR-HdB, 50.1 IT-Sicherheit, Rn. 159; *Böhm/Lolagar*, Int. Cybersecur. Law Rev. (2021) 2:317–337 (318).

[11] Hassan, Nihad A. & Hijazi, Rami, Open Source Intelligence Methods and Tools. A Practical Guide to Online Intelligence, Apress, Berkeley, CA. 2018.

Navigationsdaten, Umweltdaten) genannt werden.[12] Technische Unterstützung bei solchen Open Source-Ermittlungen kann den Ermittler*innen beispielsweise eine Software bieten, die als Browser-Erweiterung installiert wird und relevante Daten wie E-Mail-Adressen oder Informationen über Bilder einer aufgerufen Webseite erfasst. Wenn im Rahmen von Open Source-Ermittlungen jedoch größere Datenmengen verarbeitet werden sollen, werden hierfür zumeist sogenannte „Web-Crawler" (auch „Web-Scraper") eingesetzt.[13] Diese sind in der Lage, öffentlich zugängliche Informationen aus dem Internet automatisiert zu erfassen und in großen Datenbanken zu persistieren.[14] Die gecrawlten Daten können anschließend – abhängig vom jeweiligen Auswertungszweck – anhand verschiedener Analysemethoden aufbereitet, miteinander verknüpft und auf ermittlungsrelevante Informationen durchsucht werden.[15] Auf diese Weise lassen sich Erkenntnisse generieren, die über den Informationsgehalt und die Aussagekraft der einzelnen erhobenen Datensätze weit hinausgehen.[16] Solche Erkenntnisse nehmen bei strafrechtlichen Ermittlungen im Darknet einen besonderen Stellenwert ein, da die meisten „herkömmlichen" Ermittlungsansätze, die zur Aufklärung von Straftaten in nicht-anonymen Teilbereichen des Internets zur Anwendung kommen, im Darknet nicht zielführend sind. Der Grund hierfür ist in der technischen Infrastruktur zu sehen, auf der Darknet-Technologien basieren. Diese wird nachfolgend in Kürze erläutert.

2.2 Darknet-Technologien zur anonymen Nutzung des Internets

Der Begriff Darknet[17] bezeichnet denjenigen Teilbereich des Internets, der größtmögliche Anonymität bei der Nutzung des Internets gewährleistet, indem IP-Adressen verschleiert und Kommunikationsinhalte verschlüsselt werden. Infolgedessen ist es für Dritte nicht mehr nachvollziehbar, wer mit wem über was im Internet kommuniziert. Um dies zu ermöglichen, bilden Darknet-Technologien oberhalb der existierenden technischen Infrastruktur des Internets eigene Kommunikationsräume, die als „Overlay-Netzwerke" bezeichnet werden. Es existieren mehrere solcher „Overlay-Netzwerke" wie zum Beispiel Tor, I2P oder Freenet,[18] die jeweils technisch unabhängig voneinander betrieben werden. Gemeinsam ist den genannten Technologien, dass sie ihren Nutzer*innen die Möglichkeit

[12] Hassan, Nihad A. & Hijazi, Rami, Open Source Intelligence Methods and Tools. A Practical Guide to Online Intelligence, Apress, Berkeley, CA. 2018.

[13] *Rückert*, Mit künstlicher Intelligenz auf Verbrecherjagd: Einsatz von Gesichtserkennungstechnologie zur Aufklärung der „Kapitolverbrechen", VerfBlog vom 22.01.2021, abrufbar unter https://verfassungsblog.de/ki-verbrecherjagd (letzter Zugriff: 30.01.2022).

[14] *Rückert*, Mit künstlicher Intelligenz auf Verbrecherjagd: Einsatz von Gesichtserkennungstechnologie zur Aufklärung der „Kapitolverbrechen", VerfBlog vom 22.01.2021, abrufbar unter https://verfassungsblog.de/ki-verbrecherjagd (letzter Zugriff: 30.01.2022).

[15] *Kalpakis* et al., OSINT and the Dark Web, in: Open Source Intelligence Investigation, S. 111–132.

[16] *Rückert*, ZStW 2017, 129 (2); 302 (328).

[17] Zur berechtigten Kritik am Präfix „dark" und den damit einhergehenden negativen Assoziations- und Deutungsrahmen vgl. *Bovermann*, Framing-Check: Darknet, Süddeutsche Zeitung (03.05.2019), abrufbar unter https://www.sueddeutsche.de/kultur/framing-darknet-tor-anonym-internet-silk-road-1.4367011 (letzter Zugriff: 30.01.2022).

[18] *Kalpakis* et al, OSINT and the Dark Web, in: Open Source Intelligence Investigation, S. 111–132.

bieten, sich anonym im Internet zu bewegen und das Hinterlassen von (Daten-)Spuren weitgehend zu minimieren. Dies kann in der Praxis aus zahlreichen legitimen Gründen wünschenswert sein, etwa um die eigene Privatsphäre zu schützen und das „Tracking" sensibler Daten zu unterbinden. Durch die Nutzung von Darknet-Technologien kann außerdem staatliche Informationskontrolle und Internetzensur in autoritären Regimen umgangen werden.[19] Hinzu kommt, dass sogenannte „chilling effects", welche die Nichtausübung von Grund- und Freiheitsrechten aus Furcht vor Überwachung bezeichnen, durch den Einsatz von Darknet-Technologien reduziert werden können.[20] Das Verbergen der eigenen Identität bei der Veröffentlichung von Informationen und Dokumenten erleichtert außerdem die Arbeit von Journalist*innen, Menschenrechtsaktivist*innen und Whistleblower*innen weltweit. Ein prominentes Beispiel hierfür ist die Gruppe „Raqqa is being slaughtered silently", welche Darknet-Technologien benutzte, um auf Menschenrechtsverletzungen des sogenannten „Islamischen Staates" in Syrien aufmerksam zu machen.[21] Ein weiteres bekanntes Beispiel stellt die Veröffentlichung der „Panama Papers" dar, die der Süddeutschen Zeitung anonym zugespielt wurden.[22] Darknet-Technologien haben demnach zahlreiche „helle" und gesellschaftlich wünschenswerte Verwendungszwecke vorzuweisen, die in der medialen Berichterstattung allerdings nur wenig Beachtung finden.

3 Aktuelle Kriminalitätsphänomene im Darknet

Oftmals werden Darknet-Technologien von ihren Nutzer*innen jedoch auch dafür verwendet, Straftaten im Internet möglichst anonym zu begehen. Für die Strafverfolgungspraxis geht damit einher, dass aufgrund der Anonymisierung von IP-Adressen letztere nicht mehr als Anhaltspunkte für die Identifizierung von tatverdächtigen Internetnutzer*innen herangezogen werden können. Da IP-Adressen im Rahmen von „klassischen" Internetermittlungen einen zentralen, äußerst wichtigen Ermittlungsansatz darstellen, wird die Aufklärung von Straftaten, die mittels Darknet-Technologien begangen werden, in der Praxis erheblich erschwert. Aus diesem Grund haben sich in anonymen Teilbereichen des Internets eine Vielzahl verschiedener Kriminalitätsphänomene etabliert, die in ihrer Gesamtheit als darknetspezifische

[19] Über die Darknet-Technologie „Tor" kann teilweise auf westliche Internetdienste zugegriffen werden, deren Inhalte in Ländern wie China oder Saudi-Arabien blockiert werden. Zur rechtlichen Stellung des Tor-Darknets im Ausland und generell zu dessen „hellen Seiten" vgl. *Watson*, Washington University Global Studies Law Review 2012, Vol. 11, Issue 3, S. 715–737.

[20] *Bartl/Moßbrucker/Rückert*, Angriff auf die Anonymität im Internet, S. 18. Dort wird der Begriff der „chilling effects", welcher der anglo-amerikanischen Rechtstradition entspringt, näher erläutert.

[21] Vgl. *Mihr*, Das Darknet als geschützter Raum gegen Überwachung und Selbstzensur, bpb (07.03.2017), abrufbar unter https://www.bpb.de/dialog/netzdebatte/239003/das-darknet-als-geschuetzter-raum-gegen-ueberwachung-und-selbstzensur (letzter Zugriff: 30.01.2022).

[22] Ausführlich zum Leak der sogenannten „Panama Papers" siehe *Süddeutsche Zeitung* (Erscheinungsdatum unbekannt), Das sind die Panama Papers, abrufbar unter https://panamapapers.sueddeutsche.de/articles/56ff9a28a1bb8d3c3495ae13 (letzter Zugriff 29.04.2022).

Kriminalität bezeichnet werden können. Letztere spielt sich häufig auf Webseiten ab, die ausschließlich dafür genutzt werden, illegale Handels- und Tauschgeschäfte anonym anbahnen und abwickeln zu können. Diese Webseiten lassen sich aufgrund ihrer unterschiedlichen inhaltlichen Ausrichtung und Geschäftsmodelle in Handelsplattformen (auch „Kryptomärkte" genannt, siehe *Abbildung 1*), Diskussionsforen (siehe *Abbildung 2*) und pädokriminelle Tauschbörsen (siehe *Abbildung 3*) untergliedern. Das Spektrum der über diese Webseiten verwirklichten Straftaten ist äußerst vielschichtig und umfasst den Handel mit Betäubungs- und Arzneimitteln i.S.d. §§ 29 ff. BtMG und §§ 95 f. AMG, Verstöße gegen die §§ 51, 52 WaffG, die Verbreitung kinderpornografischer Inhalte i.S.d. § 184b StGB sowie Geld- und Urkundenfälschungsdelikte i.S.d. §§ 146 ff. und §§ 267 ff. StGB. Des Weiteren wird mit verschiedenen Formen von Schadsoftware i.S.d. § 202c Abs. 1 Nr. 2 und § 303a Abs. 3 StGB, ausgespähten Daten i.S.v. § 202d StGB sowie Geldwäschedienstleistungen i.S.v. § 261 StGB gehandelt. Darüber hinaus werden aus Cyber-Angriffen stammende personenbezogene sowie betriebsinterne Daten über sogenannte „Doxing"-Webseiten beziehungsweise „Leaking"-Plattformen veröffentlicht und Botnetze über anonyme „Command & Control-Server" gesteuert.[23] Seit dem Aufkommen der COVID-19-Pandemie wird zudem mit gefälschten Impfpässen und Impfetiketten gehandelt, wodurch seit der Einführung des § 277 StGB im November 2021 ein eigenständiger Straftatbestand verwirklicht wird.

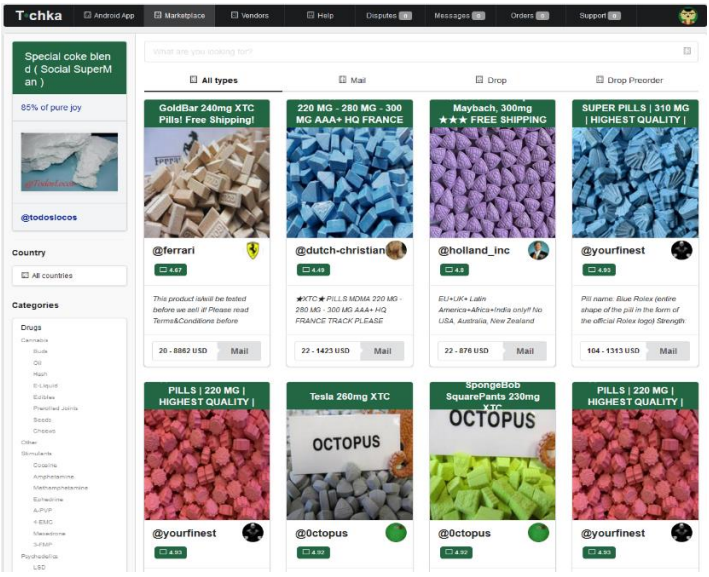


Abbildung 1: Angebote für Ecstasy-Tabletten auf dem Kryptomarkt „Tochka"

[23] *Stam*, ZIS 2017, 547 (547).

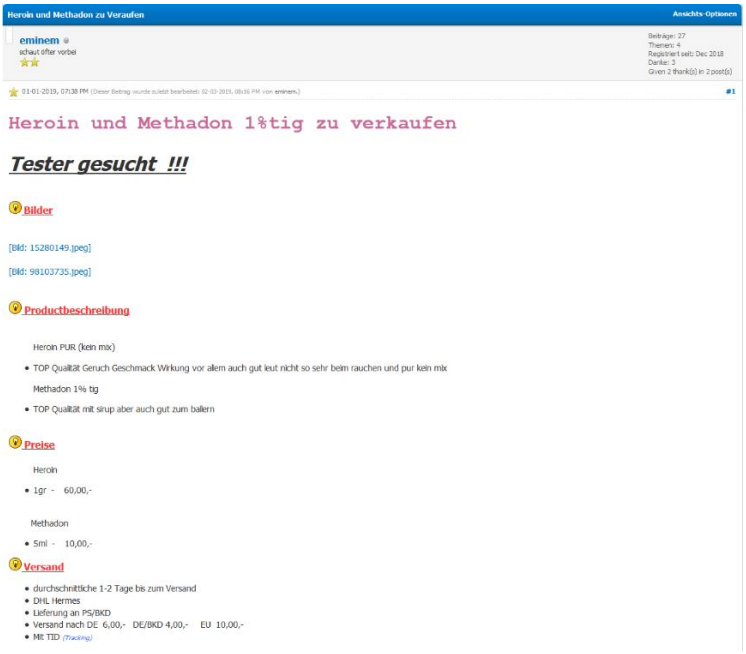


Abbildung 2: Angebotsbeschreibung (Heroin, Methadon) auf dem Diskussionsforum „DiDW 3.0“

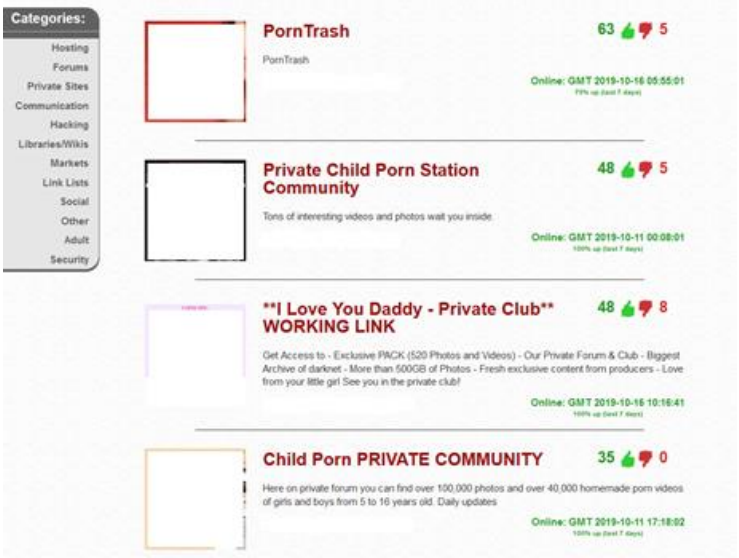


Abbildung 3: Pädokriminelle Tauschbörsen mit kinderpornografischen Inhalten

4 Darknetspezifische Open Source-Ermittlungen

Zur Aufklärung dieser Straftaten kommt aus technischer Perspektive die Nutzung von Open Source Intelligence (OSINT) in Betracht. Als wohl bekanntestes Praxisbeispiel für den Einsatz von OSINT zur Strafverfolgung im Darknet kann die Identifizierung des Betreibers der Handelsplattform „Silk Road" genannt werden, die zwischen 2011 und 2013 den Markt für Drogen im Darknet dominierte.[24] Die Plattform konnte vom FBI abgeschaltet werden, nachdem eine Sammlung und Auswertung von frei zugänglichen Datenbeständen aus dem Internet US-amerikanische Ermittler*innen auf die Spur des Plattform-Administrators von „Silk Road" brachten. Dieser hatte im Januar 2011, also zu einem Zeitpunkt, in dem seine Plattform der Öffentlichkeit noch nicht bekannt war, unter dem Nicknamen „altoid" einen Werbepost für seinen Online-Marktplatz „Silk Road" veröffentlicht.[25] Die Ermittler*innen suchten auf öffentlichen Webseiten nach weiteren unter diesem Pseudonym veröffentlichten Posts und konnten schließlich eine Nachricht in einem Bitcoin-Forum ausfindig machen, in dem „altoid" als Kontaktmöglichkeit eine E-Mail-Adresse angegeben hatte, die sich aus seinem bürgerlichen Vor- und Nachnamen zusammensetzte.[26] Daraufhin konnten die Ermittler*innen ihre OSINT-Recherchen auf einen damals 29-jährigen US-Amerikaner eingrenzen, der in diversen sozialen Medien teilweise gleichlautende Aussagen gepostet hatte, wie der gesuchte Plattform-Administrator von „Silk Road" im Darknet. Diese und weitere Erkenntnisse führten schlussendlich dazu, dass der Betreiber von „Silk Road" in der Offline-Welt identifiziert und von Beamt*innen des FBI festgenommen werden konnte.[27]

4.1 Grundrechtseingriff durch Open Source-Ermittlungen

Seitdem die Zentralstelle Cybercrime Bayern (ZCB) im Juli 2020 mit der Erprobung des „Dark Web Monitors" begonnen hat, sind Ermittlungen dieser Art grundsätzlich auch in Deutschland denkbar. In welchen Grenzen öffentlich zugängliche Datenbestände von staatlichen Stellen in strafprozessualen Ermittlungsverfahren gesammelt, gespeichert und sodann zu Strafverfolgungszwecken ausgewertet werden dürfen, ist eine in der

[24] Ausführlich zur Handelsplattform „Silk Road" siehe *Tzanetakis*, Zum Phänomen der Drogenmärkte im Darknet, in: *Tzanetakis/Stöver* (Hrsg.), Drogen, Darknet und Organisierte Kriminalität, S. 11.

[25] Vgl. *Van Note*, How OSINT Took Down the Dark Web's Silk Road, Scopenow (22.07.2020), abrufbar unter https://www.skopenow.com/news/how-osint-took-down-the-dark-webs-silk-road (letzter Zugriff: 30.01.2022).

[26] Vgl. *Van Note*, How OSINT Took Down the Dark Web's Silk Road, Scopenow (22.07.2020), abrufbar unter https://www.skopenow.com/news/how-osint-took-down-the-dark-webs-silk-road (letzter Zugriff: 30.01.2022). Der betreffende Post vom 11.10.2011 ist auf „Bitcointalk" bis heute öffentlich einsehbar, vgl. https://bitcointalk.org/index.php?action=profile;u=3905;sa=showPosts (letzter Zugriff: 30.01.2022).

[27] *Tanriverdi*, Mutmaßlicher Betreiber der Drogen-Plattform „Silk Road": Drogenhandel, Mordversuch – und den Klarnamen angegeben, Süddeutsche Zeitung (04.10.2013), abrufbar unter https://www.sueddeutsche.de/digital/mutmasslicher-betreiber-der-drogen-plattform-silk-road-drogenhandel-mordversuch-und-den-klarnamen-angegeben-1.1786870 (letzter Zugriff: 30.01.2022).

Rechtswissenschaft bislang nur spärlich diskutierte Frage.[28] Open Source-Ermittlungen bedürfen jedenfalls dann einer strafprozessualen Ermächtigungsgrundlage, wenn durch sie in das Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG eingegriffen wird. Letzteres schützt die Befugnis einer Person, grundsätzlich selbst über die Preisgabe und Verwendung ihrer persönlichen Daten zu bestimmen.[29] Laut BVerfG stellt die Erhebung allgemein zugänglicher, personenbezogener Daten grundsätzlich keinen Grundrechtseingriff dar, da eine Kenntnisnahme solcher Informationen dem Staat grundsätzlich nicht verwehrt ist.[30] Der grundrechtliche Schutz des Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG greift allerdings dann, wenn personenbezogene Informationen von staatlichen Stellen in einer Art und Weise genutzt und miteinander verknüpft werden, die die Betroffenen weder überschauen noch beherrschen können.[31] Dabei kommt es aufgrund der modernen Möglichkeiten der Datenverarbeitung nicht mehr auf einen eindeutigen Personenbezug der ausgewerteten Daten an.[32] Vielmehr werden auch anonymisierte oder pseudonymisierte Informationen von Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG geschützt, wenn eine De-Anonymisierung oder De-Pseudonymisierung der jeweiligen Daten durch staatliche Stellen möglich ist.[33] Sollten frei zugängliche Informationen aus dem Internet von staatlichen Stellen gezielt zusammengetragen und zu Strafverfolgungszwecken ausgewertet werden, bedarf es für entsprechende Ermittlungshandlungen daher einer strafprozessualen Ermächtigungsgrundlage.[34]

4.2 Ermächtigungsgrundlagen der §§ 98a, 98c StPO

Auch wenn dies auf den ersten Blick nahe liegt, kommt die Vorschrift des § 98a StPO, welche die strafprozessuale Rasterfahndung regelt, nicht als Ermächtigungsgrundlage für die automatisierte Sammlung und Auswertung von öffentlich zugänglichen Informationen aus dem Internet in Betracht.[35] Zwar besteht eine gewisse „Verwandtschaft" zwischen Open Source-Ermittlungen und der Rasterfahndung aus § 98a StPO darin, dass im Rahmen beider Verfahren große Datensätze zu Ermittlungszwecken maschinell miteinander abgeglichen werden.[36] Allerdings werden im Rahmen von § 98a StPO bereits bestehende, nicht-öffentliche Datensätze ausgewertet, während bei softwaregestützten Open Source-Ermittlungen in aller Regel Informationen zum Einsatz kommen, die öffentlich zugänglich sind, dafür aber zunächst noch mittels Web-Crawling erhoben werden müssen. Gleiches

[28] Ein erster, ausführlicher Beitrag zur Zulässigkeit der Verarbeitung öffentlich zugänglicher Daten im strafrechtlichen Ermittlungsverfahren findet sich bei *Rückert*, ZStW 2017, 129 (2); 302–333.
[29] BVerfG, Urteil des Ersten Senats vom 15. Dezember 1983 - 1 BvR 209/83, Rn. 147.
[30] BVerfG, Urteil vom 27. 2. 2008 - 1 BvR 370/07, 1 BvR 595/07, Rn. 308.
[31] BVerfG, Beschluss des Ersten Senats vom 13. Juni 2007 - 1 BvR 1550/03, Rn. 87.
[32] *Rückert*, ZStW 2017, 129 (2); 302 (313–314).
[33] *Rückert*, ZStW 2017, 129 (2); 302 (314).
[34] Siehe hierzu auch *Golla*, DSRITB 2018, 871 (871–872).
[35] Vgl. *Rückert*, ZStW 2017, 129 (2); 302 (316).
[36] Vgl. *Rückert*, ZStW 2017, 129 (2); 302 (316).

gilt in Hinblick auf die Vorschrift des § 98c StPO, welche den maschinellen Abgleich von bereits erhobenen, justizinternen Informationen regelt.[37]

4.3 Rückgriff auf die Vorschriften der §§ 161, 163 StPO

Da keine weiteren strafprozessualen Spezialbefugnisse ersichtlich sind, die die Vornahme von Open Source-Ermittlungen rechtfertigen könnten, kommt als Rechtsgrundlage für entsprechende Ermittlungshandlungen nur ein Rückgriff auf die Ermittlungsgeneralklausel der §§ 161, 163 StPO in Betracht. Ob diese als taugliche Rechtsgrundlage für Open Source-Ermittlungen herangezogen werden kann, hängt wiederum von der Intensität der Grundrechtseingriffe ab, die von entsprechenden Ermittlungshandlungen in Hinblick auf das Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG verwirklicht werden. Hinsichtlich automatisierter Open Source-Ermittlungen kann festgestellt werden, dass sich diese in ihrer Grundrechtsintensität grundlegend von der „manuellen" Datenerhebung und Auswertung durch menschliche Ermittler*innen unterscheiden.[38] Sie sind mit Blick auf Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG wesentlich eingriffsintensiver, da softwaregestützte Datenerhebungs- und Analysemöglichkeiten im Vergleich zu menschlichen Open Source-Recherchen personell und zeitlich nicht beschränkt sind. Hierin liegt der große praktische Nutzen von softwaregestützten Open Source-Ermittlungen im Vergleich zu händischen Recherchen durch einzelne Ermittlungsbeamt*innen, die nicht selten einer „Suche nach der Nadel im Heuhaufen" gleichen. Je effektiver die automatisierte Sammlung und Auswertung von öffentlich zugänglichen Daten zu Strafverfolgungszwecken funktioniert, desto höher ist folglich auch die Intensität der Grundrechtseingriffe in Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG zu bewerten, die durch entsprechende Ermittlungshandlungen verwirklicht werden.

4.4 Vorliegen eines Anfangsverdachts

Diese Erkenntnis wirkt sich wiederum auf die Frage aus, ob Open Source-Ermittlungen noch auf die Ermittlungsgeneralklausel der §§ 161, 163 StPO gestützt werden können. Damit letztere überhaupt als Rechtsgrundlage für die automatisierte Sammlung und Auswertung von öffentlich zugänglichen Daten herangezogen werden kann, müssen zunächst einmal „zureichende tatsächliche Anhaltspunkte" dafür vorliegen, dass eine verfolgbare Straftat i.S.d. § 152 Abs. 2 StPO begangen wurde und die ergriffenen Ermittlungsmaßnahmen der Sachverhaltserforschung i.S.d. § 160 Abs. 1 StPO dienen. Vorfeldermittlungen und sogenannte „Verdächtigengewinnungseingriffe" sind demnach unzulässig, da ansonsten die Begrenzungsfunktion des Anfangsverdachts zum Schutze vor

[37] *Rückert*, ZStW 2017, 129 (2); 302 (318).
[38] Ebenso *Rückert*, ZStW 2017, 129 (2); 302 (326).

unberechtigter Strafverfolgung ausgehebelt würde.[39] Eine verdachtslose Erhebung und Auswertung von allgemein zugänglichen Daten im Rahmen von Open Source-Ermittlungen ohne tatsächliche Anhaltspunkte für das Vorliegen einer strafbaren Handlung kann daher nicht auf die Vorschriften der §§ 161, 163 StPO gestützt werden. In der Praxis folgt daraus, dass lediglich solche Webseiten zu Strafverfolgungszwecken gecrawlt werden dürfen, die eine (eindeutige) kriminelle Ausrichtung aufweisen. Dies ist beispielsweise bei im Darknet angesiedelten, illegalen Online-Marktplätzen der Fall, die ausdrücklich auf den Handel von inkriminierten Waren und Dienstleistungen ausgerichtet sind.[40] Auf im Darknet betriebenen Diskussionsforen, die im Gegensatz zu illegalen Online-Marktplätzen auch als Treffpunkte für den anonymen Austausch über Nachrichten aus Politik und Wirtschaft, IT-Sicherheit und andere strafrechtlich irrelevante Themen genutzt werden, ist eine Datenerhebung und -Auswertung mittels Web-Crawling jedoch unzulässig. Da der Anwendungsbereich der §§ 161, 163 StPO in diesen Konstellationen mangels Anfangsverdachts nicht eröffnet ist, fehlt es in letztgenannten Fällen an einer tauglichen Rechtsgrundlage, sodass entsprechende Ermittlungshandlungen unterbleiben müssen.[41]

4.5 „Monitoring“ mit anonymisierten Daten

Ausnahmsweise kommt eine automatisierte Erhebung und Auswertung von öffentlich zugänglichen Daten allerdings auch in Fällen in Betracht, in denen die gecrawlten Webseiten keinerlei Verbindungen zur kriminellen Szene im Darknet aufweisen. Zwar kann eine Sammlung und Analyse von öffentlich zugänglichen Daten in diesen Fällen mangels Anfangsverdacht nicht auf die Ermittlungsgeneralklausel aus §§ 161, 163 StPO gestützt werden. Solange die gesammelten Daten von staatlichen Stellen nicht de-anonymisiert oder de-pseudonymisiert werden können, ist dies mangels Grundrechtseingriff in Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG jedoch auch nicht erforderlich. Aus technischer Sicht ist es nämlich möglich, öffentlich zugängliche Daten zu erheben und in Datenbanken zu persistieren, die vollständig anonymisiert wurden und daher keinerlei Personenbezug mehr aufweisen.[42] In der Praxis kann dies vor allem in Konstellationen sinnvoll sein, in denen es um das reine „Monitoring“ von einzelnen Webseiten im Darknet geht, also zum Beispiel um die Frage, welche Gesprächsthemen in einem bestimmten Online-Forum diskutiert werden oder welche Entwicklungen und Trends sich in Hinblick auf bestimmte, über illegale Online-Marktplätze gehandelte

[39] Siehe hierzu bereits *Bauer*, Soziale Netzwerke und strafprozessuale Ermittlungen, S. 141.
[40] Der Betrieb solcher Plattformen ist seit Oktober 2021 eigenständig unter Strafe gestellt (§ 127 StGB).
[41] Aus technischer Perspektive besteht in einigen dieser Fällen allerdings die Möglichkeit, das Crawling auf bestimmte, ausschließlich kriminelle Teilbereiche eines im Darknet betriebenen Diskussionsforums (wie zum Beispiel einen strafrechtlich relevanten „Thread“ über den Verkauf von Waffen) zu begrenzen.
[42] Murthy, S. et al.: A comparative study of data anonymization techniques. In (IEEE, Hrsg.) 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). S. 206-309, 2019.

Betäubungsmittel oder verwendete Zahlungsmittel abzeichnen. Die Einholung dieser Informationen kann auch ohne Personenbezug für die Ermittler*innen von Bedeutung sein, da eine eingehende Kenntnis der Szene und der auf den Plattformen im Darknet üblichen Abläufe hilfreich ist, um eine passende Strategie für mögliche weitere Ermittlungsansätze zu erarbeiten.[43]

4.6 Unions- und datenschutzrechtliche Vorgaben

Im Rahmen des „Monitorings" mit vollständig anonymisierten Datensätzen können einzelne tatverdächtige Personen – wie es im Falle des Betreibers der Handelsplattform „Silk Road" geschehen ist – allerdings nicht identifiziert werden. Um dies zu ermöglichen, müssen von den Ermittler*innen notwendigerweise auch solche Informationen erhoben und ausgewertet werden, die einen Personenbezug aufweisen oder von staatlichen Stellen zumindest wieder de-anonymisiert beziehungsweise de-pseudonymisiert werden können. In diesen Fällen ist neben dem Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG weiterhin erforderlich, dass unionsrechtliche Vorgaben beachtet werden, da Open Source-Ermittlungen andernfalls gegen Unionsgrundrechte (Art. 7 und 8 der GRCh) verstoßen würden. Um eine Verletzung von unionsrechtlichen Bestimmungen vorzubeugen, ist die Einhaltung der Richtlinie (EU) 2016/680 des Europäischen Parlaments und des Rates vom 27. April 2016 zum Schutz natürlicher Personen bei der Verarbeitung personenbezogener Daten durch die zuständigen Behörden zum Zwecke der Verhütung, Ermittlung, Aufdeckung oder Verfolgung von Straftaten oder der Strafvollstreckung (kurz EU-RL 2016/680) maßgeblich.[44] Gem. Art. 4 der EU-RL 2016/680 dürfen personenbezogene Daten, die für Strafverfolgungszwecke erhoben wurden, von den Ermittler*innen nicht in einer mit diesem Zweck nicht zu vereinbarenden Weise verarbeitet werden. Die betreffenden Daten müssen außerdem in einer Weise verarbeitet werden, die eine angemessene Sicherheit gewährleistet, einschließlich des Schutzes vor unbefugter oder unrechtmäßiger Verarbeitung und vor unbeabsichtigtem Verlust, unbeabsichtigter Zerstörung oder unbeabsichtigter Schädigung. Gem. Art. 20 EU-RL 2016/680 muss die Speicherung personenbezogener Daten zudem im Sinne der Datenminimierung verhältnismäßig sein. Weiterhin setzt Art. 27 EU-RL 2016/680 für Datenverarbeitungsvorgänge, im Rahmen derer „neue Technologien" verwendet werden, die vorherige Vornahme einer Abschätzung der sich hieraus ergebenen Folgen für die im Einzelnen betroffenen Personen (sogenannte „Datenschutz-Folgenabschätzung", vgl. § 67 BDSG) voraus.

---

[43] Siehe hierzu etwa *Krause*, NJW 2018, 678 (679–680).

[44] In Deutschland umgesetzt wurden die Vorgaben der EU-RL 2016/680 durch das Gesetz zur Umsetzung der RL (EU) 2016/680 im Strafverfahren sowie zur Anpassung datenschutzrechtlicher Bestimmungen an die Verordnung (EU) 2016/679 vom 20.11.2019, vgl. BGBl. 2019 I, S. 1724 ff.

4.7 Persönlichkeitsrelevanz der verarbeiteten Daten

Ist die softwaregestützte Sammlung und Auswertung von öffentlich zugänglichen Datenbeständen mit unions- und datenschutzrechtlichen Bestimmungen vereinbar und liegt ein Anfangsverdacht i.S.d. §§ 161, 163 StPO vor, hängt die Zulässigkeit von Open Source-Ermittlungen weiterhin von dem Inhalt und der Art derjenigen Daten ab, die von den Ermittler*innen gesammelt und anschließend zu Strafverfolgungszwecken ausgewertet werden sollen. Beschränkt sich die Erhebung, Speicherung und Auswertung auf inhaltlich eng begrenzte Daten, die weder höchstpersönliche Informationen umfassen, noch die Erstellung von Persönlichkeits- oder Bewegungsprofilen der Verdächtigen ermöglichen, können Open Source-Ermittlungen grundsätzlich auf die Ermittlungsgeneralklausel der §§ 161, 163 StPO gestützt werden, da die hiermit verbundenen Grundrechtseingriffe in das Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG als geringfügig anzusehen sind. Als Beispiel hierfür kann etwa die softwaregestützte Sammlung und Auswertung von sämtlichen, auf einem bestimmten illegalen Online-Marktplatz im Darknet veröffentlichten Verkaufsangeboten genannt werden. Mithilfe solcher Open Source-Ermittlungen können Strafverfolgungsbehörden beispielsweise diejenigen Personen identifizieren, die auf den gecrawlten Webseiten über besonders hohe Handels- beziehungsweise Umsatzvolumina verfügen. Gegenüber diesen, in der Szene häufig als „Power-Seller*innen“ oder „Top-Vendor*innen“ bezeichneten Tatverdächtigen können sodann weitergehende Ermittlungsmaßnahmen – wie etwa die Durchführung von Testkäufen oder der Einsatz von verdeckten Ermittler*innen – eingeleitet werden. Auf diese Weise wäre es möglich, mithilfe von OSINT die ohnehin knappen personellen Ressourcen der Strafverfolgungsbehörden zu priorisieren und vorhandene Ermittlungskapazitäten möglichst effektiv zur Identifizierung von einzelnen tatverdächtigen Personen einzusetzen. Werden im Rahmen von Open Source-Ermittlungen im Gegensatz zu dem beschriebenen Praxisbeispiel jedoch Daten gesammelt und ausgewertet, die höchstpersönliche Informationen umfassen und daher die Erstellung von umfangreichen Persönlichkeits- oder Bewegungsprofilen einzelner tatverdächtiger Personen ermöglichen, kann nicht mehr von einer „bloß geringfügigen“ Beeinträchtigung des Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG ausgegangen werden. Als Beispiel für einen solchen, nicht mehr bloß geringfügigen Grundrechtseingriff kann die Erhebung und Auswertung von öffentlich zugänglichen Daten herangezogen werden, die eine hohe Persönlichkeitsrelevanz aufweisen und daher besonders „sensibel“ sind. Hierunter fallen zum Beispiel Daten, aus denen – in Anlehnung an Art. 10 der EU-RL 2016/680 – die ethnische Herkunft, sexuelle Orientierung, politische Meinung sowie religiöse oder weltanschauliche Überzeugung einer tatverdächtigen Person hervorgehen.[45] Werden derlei Informationen im Rahmen von Open Source-Ermittlungen automatisiert erhoben und im Anschluss zu Strafverfolgungszwecken ausgewertet, ist davon auszugehen, dass

[45] Weitere Beispiele für in Art. 10 der EU-RL 2016/680 genannte „sensible“ Daten sind genetische Daten, biometrische Daten, Gesundheitsdaten sowie Daten zum Sexualleben oder zur Gewerkschaftszugehörigkeit einer Person.

die hierdurch verwirklichten Grundrechtseingriffe in das Recht auf informationelle Selbstbestimmung mehr als nur geringfügig sind, sodass entsprechende Ermittlungshandlungen nicht auf die Ermittlungsgeneralklausel der §§ 161, 163 StPO gestützt werden können und demnach unterbleiben müssen.

4.8 Datenmenge und Anzahl an ausgewerteten Datenquellen

Neben der Persönlichkeitsrelevanz der gecrawlten Daten stellt die Menge an gesammelten Datensätzen sowie die Anzahl der ausgewerteten Datenquellen ein weiteres Kriterium für die Beurteilung der Grundrechtsintensität von Open Source-Ermittlungen dar. Denn je mehr Daten softwaregestützt gesammelt werden und je diverser die Datenquellen sind, aus denen die betreffenden Informationen erhoben werden, desto aufschlussreicher sind in aller Regel auch die Erkenntnisse, die aus einer Zusammenschau dieser Daten resultieren. Selbst wenn die ausgewerteten Daten keine höchstpersönlichen Informationen umfassen, können auf diese Weise Beziehungen oder Zusammenhänge zwischen einzelnen tatverdächtigen Personen, Personengruppierungen, Institutionen, Organisationen, Objekten und Sachen ermittelt werden, die in Hinblick auf Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG für einen nicht mehr bloß geringfügigen Grundrechtseingriff sprechen. Die automatisierte Erhebung und Auswertung von Daten, die allesamt von ein und derselben Webseite stammen, ist somit weniger eingriffsintensiv, als wenn diese Informationen – wie im Beispielsfall der Handelsplattform „Silk Road“ – anschließend noch mit weiteren, externen Datenquellen abgeglichen werden. Je größer die Menge an gesammelten Datensätzen und je diverser die Anzahl der ausgewerteten Datenquellen ist, desto intensiver wird folglich im Rahmen von Open Source-Ermittlungen in das Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 1 i.V.m. Art 1 Abs. 1 GG eingegriffen. In diesen Konstellationen kann demnach nicht mehr davon ausgegangen werden, dass Ermittlungshandlungen bloß geringfügig in das Recht auf informationelle Selbstbestimmung der Betroffenen eingreifen. Dementsprechend können Open Source-Ermittlungen in den beschriebenen Konstellationen auch nicht mehr auf die Ermittlungsgeneralklausel aus §§ 161, 163 StPO gestützt werden und müssen in der Praxis unterbleiben.

4.9 Grad der algorithmischen Datenverarbeitung

Neben qualitativen und quantitativen Aspekten der Datenverarbeitung ist die Frage, ob automatisierte Open Source-Ermittlungen auf die Vorschriften der §§ 161, 163 StPO gestützt werden können, weiterhin von der Art der algorithmischen Verarbeitung abhängig, mit dem die erhobenen Datenbestände ausgewertet werden. Wenn öffentlich zugängliche Datenbestände von einem Webcrawler gesammelt und in Datenbanken persistiert werden, damit sie anschließend nach bestimmten Schlüsselbegriffen („Keywords“) durchsucht werden können, basiert die maschinelle Entscheidungsfindung auf sogenannten „String-Matching-Algorithmen“ und ist für Dritte weitgehend erklärbar beziehungsweise nachvollziehbar. Kommen im Rahmen von Open Source-Ermittlungen

jedoch Algorithmen auf Basis von künstlicher Intelligenz (KI) zum Einsatz, ist die mit der Datenanalyse einhergehende maschinelle Entscheidungsfindung wesentlich komplexer und von den Ermittler*innen teilweise nicht mehr rekonstruierbar, sodass sie – wenn überhaupt – nur eingeschränkt validiert beziehungsweise verifiziert werden kann.[46] Als Praxisbeispiel für eine solche besonders grundrechtsintensive KI-basierte Datenanalyse kann etwa die Sammlung und Auswertung verschiedener Postings ein und derselben, bislang unbekannten Person in einem Diskussions-Forum genannt werden, anhand derer ermittelt werden soll, welche soziodemografischen Variablen – wie zum Beispiel Alter, Geschlecht, Bildungsgrad oder ethnische Herkunft – dieser Person zugeordnet werden können (sogenanntes „authorship profiling“).[47] In diesen Konstellationen ist es nicht die Qualität oder Quantität der gesammelten Daten selbst, sondern vielmehr die KI-basierte Analyse dergleichen, die eine Erstellung von umfangreichen Persönlichkeits- und Bewegungsprofilen der gesuchten tatverdächtigen Person möglich macht. Aufgrund der damit einhergehenden Eingriffsintensität kann in den beschriebenen Fällen nicht mehr von einer bloß geringfügigen Beeinträchtigung des Art. 2 Abs. 1 i.V.m. Art 1 Abs. 1 GG ausgegangen werden, sodass die Ermittlungsgeneralklausel aus §§ 161, 163 StPO nicht als Rechtfertigung für entsprechende Open Source-Ermittlungen herangezogen werden kann. Der Einsatz von KI-basierten Datenanalysen wie dem „authorship profiling“ muss in der Praxis daher mangels Rechtsgrundlage unterbleiben.

4.10 Grad der algorithmischen Datenverarbeitung

Als weiteres Kriterium für die Bewertung der Eingriffsintensität von Open Source-Ermittlungen in Art. 2 Abs. 1 i.V.m. Art 1 Abs. 1 GG ist außerdem relevant, auf welchen Zeitraum sich die automatisierte Datensammlung und -Auswertung erstreckt, die zu Strafverfolgungszwecken im Darknet zum Einsatz kommt. Je weiter die gesammelten Datenbestände in die Vergangenheit zurückreichen und je größer der Zeitraum ist, in dem einzelne Personen anhand von gecrawlten Datensätzen im Rahmen von Open Source-Ermittlungen beobachtet werden, desto intensiver sind auch die Grundrechtseingriffe in das Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG zu bewerten, die durch diese Ermittlungshandlungen verwirklicht werden. Dies ergibt sich nicht zuletzt auch aus einem Vergleich mit der Vorschrift des § 163f StPO, die planmäßig angelegte Observationen von Beschuldigten regelt, die länger als 24 Stunden andauern oder an mehr als zwei Tagen stattfinden. Softwaregestützte Open Source-Ermittlungen, die darauf ausgelegt sind, die Online-Aktivitäten einzelner Personen längerfristig zu „observieren“, können demnach nicht mehr auf die Ermittlungsgeneralklausel der §§ 161, 163 StPO gestützt werden, sondern bedürfen einer eigenständigen strafprozessualen Ermächtigungsgrundlage.

---

[46] Vgl. *Staffler/Jany*, ZIS 2020, 164 (166–167).

[47] Zum KI-basierten „authorship profiling“ siehe etwa *Sierra/Montes-y-Gómez/Solorio/González*, Convolutional Neural Networks for Author Profiling, Notebook for PAN at CLEF 2017, S. 1–7.

5 Zusammenfassung und Ausblick

Zusammenfassend kann festgehalten werden, dass die Zulässigkeit von Open Source-Ermittlungen zu Strafverfolgungszwecken im Darknet nicht pauschal beurteilt werden kann, sondern je nach Art und Weise der gesammelten Daten sowie den im Einzelfall verwendeten Analyse- und Auswertungsverfahren variiert. Ob entsprechende Ermittlungshandlungen noch auf die Vorschriften der §§ 161, 163 StPO gestützt werden können, ist davon abhängig, ob durch sie mehr als bloß geringfügig in das Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 2 i.V.m. Art. 1 Abs. 1 GG eingegriffen wird. Ob dies der Fall ist, hängt wiederum von dem Vorliegen eines Anfangsverdachts i.S.d. § 152 Abs. 2 StPO, der Einhaltung von datenschutz- und unionsrechtlichen Vorgaben, der Qualität und Quantität der verwendeten Daten sowie der Art und dem Zeitraum ihrer Sammlung und Auswertung ab. Sollte eine Zusammenschau der in diesem Beitrag entwickelten Kriterien ergeben, dass Open Source-Ermittlungen mehr als bloß geringfügig in das Recht auf informationelle Selbstbestimmung aus Art. 2 Abs. 2 i.V.m. Art. 1 Abs. 1 GG eingreifen, kann die Ermittlungsgeneralklausel aus §§ 161, 163 StPO nicht mehr zur Rechtfertigung von Open Source-Ermittlungen herangezogen werden. Mangels tauglicher Rechtsgrundlage müssen entsprechende Ermittlungshandlungen in diesen Konstellationen unterbleiben. Sollen die technischen Möglichkeiten von „Big Data“ zur Strafverfolgung im Darknet in Zukunft voll ausgeschöpft werden, wird ein legislatives „Update“ der StPO daher unvermeidbar sein. Neben grundrechtlichen Erwägungen muss in diesem Zusammenhang auch über die Zulässigkeit des „Outsourcings“ von technischen Ermittlungshandlungen und die Rolle kommerzieller Anbieter*innen wie „Hunchly“[48] oder „Recorded Future“[49] diskutiert werden, die neben dem von der Zentralstelle Cybercrime Bayern (ZCB) getesteten „Dark Web Monitor“ ebenfalls Ermittlungstools für das Darknet entwickelt haben. Weiterhin sollte auch der Umstand, dass die analysierten Daten nicht validiert werden können und demnach die Gefahr von fehlerhaften Ermittlungsergebnissen in sich bergen, im Rahmen dieser Diskussion Beachtung finden. Der vorliegende Beitrag hat lediglich einen (ersten) Überblick über die bestehende Problematik gegeben. Angesichts der hohen Bedeutung dieses bislang noch wenig erschlossen Themas wären weitere Forschungsarbeiten auf dem Gebiet der Open Source-Ermittlungen jedoch wünschenswert. Feststeht allerdings bereits jetzt, dass es eine sehr komplexe Aufgabe sein wird, die existierenden technischen Möglichkeiten im Bereich der Open Source-Ermittlungen mit den hohen Anforderungen eines rechtsstaatlichen Strafverfahrens an Transparenz, Verfahrensfairness und Grundrechtsschutz in einen angemessenen Ausgleich zu bringen. Wenn der Gesetzgeber von den schnell voranschreitenden, technischen Entwicklungen im Bereich der Open

[48] Eine Online-Demo des Tools kann unter https://www.hunch.ly/osint-webinars/osint-webinar-dark-web-1252018 eingesehen werden (letzter Zugriff: 30.011.2022).
[49] Vgl. *Recorded Future*, Identify and Profile Threats From Dark Web Sources, abrufbar unter https://www.recordedfuture.com/solutions/dark-web-monitoring/ (letzter Zugriff: 30.01.2022).

Source-Ermittlungen nicht abgehängt werden will, sollte er sich daher besser zeitnah mit dieser Thematik befassen.[50]

Literaturverzeichnis

[BMR19] Bartl, M.; Moßbrucker, D.; Rückert, C.: Angriff auf die Anonymität im Internet. Folgen des sogenannten „Darknet-Paragraphen" als § 126a Strafgesetzbuch für die rechtliche Praxis – und Kollateralschäden seines Einsatzes für die Presse-, Meinungs- und Telekommunikationsfreiheiten, Reporter ohne Grenzen e.V., Berlin 2019, abrufbar unter https://www.reporter-ohne-grenzen.de/uploads/tx_lfnews/media/20190630_Darknet_Paragraf_StN-Bartl-Mossbrucker-Rueckert.pdf (letzter Zugriff: 30.01.2022).

[Ba16] Bauer, S.: Soziale Netzwerke und strafprozessuale Ermittlungen, in: Schröder, Friedrich-Christian/ Hoyer, Andreas (Hrsg.), Strafrechtliche Abhandlungen, Band 281, Duncker & Humblot Berlin, zugleich Dissertation, Bucerius Law School Hamburg 2016.

[Me20] Bayerische Staatsministerium der Justiz: Mehr Licht ins Darknet: Dark Web Monitor soll Strafverfolgungsbehörden bei Ermittlungen im Darknet verstärken, Pressemitteilung vom 27.07.2020, abrufbar unter https://www.justiz.bayern.de/presse-und-medien/pressemitteilungen/archiv/2020/69.php (letzter Zugriff: 30.01.2022).

[BL21] Böhm, I.; Lolagar, S.: Open source intelligence: Introduction, legal, and ethical considerations, Int. Cybersecur. Law Rev. (2021) 2:317–337.

[Bo19a] Borchers, D: Europäischer Polizeikongress: Weg mit dem Darknet, Heise Online (20.02.2019), abrufbar unter https://www.heise.de/newsticker/meldung/Europaeischer-Polizeikongress-Weg-mit-dem-Darknet-4313276.html (letzter Zugriff: 30.01.2022).

[Bo19b] Bovermann, P.: Framing-Check: "Darknet", Süddeutsche Zeitung (03.05.2019), abrufbar unter https://www.sueddeutsche.de/kultur/framing-darknet-tor-anonym-internet-silk-road-1.4367011 (letzter Zugriff: 30.01.2022).

[Ei20] Eisenreich, G: „Dark Web Monitor": Suchmaschine für das Darknet, DRiZ 2020, S. 290–291.

[Go18] Golla, S. J.: Lizenz zum Schürfen: Rechtsgrundlagen für die automatisierte Polizeiarbeit im Social Web, in: Taeger, Jürgen (Hrsg.), Tagungsband Herbstakademie 2018, Rechtsfragen digitaler Transformationen – Gestaltung digitaler Veränderungsprozesse durch Recht, S. 871–885, Oldenburger Verlag für Wirtschaft, Informatik und Recht.

[GR21] Grundner, B.; Rauch, M.: OEZ-Attentat in München: Als der Hass neun Menschen tötete, BR 24 vom 22.07.2021, abrufbar unter

[50] Zu diesem Ergebnis (allerdings in Hinblick auf die Nutzung von künstlicher Intelligenz und Gesichtserkennungstechnologien) siehe bereits *Rückert*, Mit künstlicher Intelligenz auf Verbrecherjagd: Einsatz von Gesichtserkennungstechnologie zur Aufklärung der „Kapitolverbrechen", VerfBlog, 2021/1/22, abrufbar unter https://verfassungsblog.de/ki-verbrecherjagd/ (letzter Zugriff: 30.01.2022).

https://www.br.de/nachrichten/bayern/oez-attentat-in-muenchen-als-der-hass-neun-menschen-toetete,SdniNXM (letzter Zugriff: 29.04.2022).

[HH18] Hassan, N. A.; Hijazi, R.: Open Source Intelligence Methods and Tools. A Practical Guide to Online Intelligence. Apress, Berkeley, CA. 2018.

[Ka16] Kalpakis, G. et.al.: OSINT and the Dark Web. In Open Source Intelligence Investigation. Springer, S. 111-132, 2016

[Kr18] Krause, B.: Ermittlungen im Darknet – Mythos und Realität, NJW 2018, 678–681.

[Mi22] Mihr, C.: Das Darknet als geschützter Raum gegen Überwachung und Selbstzensur, bpb (13.12.2016), abrufbar unter https://www.bpb.de/dialog/netzdebatte/239003/das-darknet-als-geschuetzter-raum-gegen-ueberwachung-und-selbstzensur (letzter Zugriff: 30.01.2022).

[Mu19] Murthy, Suntherasvaran et al.: A comparative study of data anonymization techniques. In (IEEE, Hrsg.) 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS). S. 206-309, 2019.

[Ob] Obermayer, B.; Obermaier F.; Wormer, V.; Jaschensky, W.: Das sind die Panama Papers, Süddeutsche Zeitung (Erscheinungsdatum unbekannt), abrufbar unter https://panamapapers.sueddeutsche.de/articles/56ff9a28a1bb8d3c3495ae13 (letzter Zugriff 29.04.2022).

[Rü22] Rückert, C.: Mit künstlicher Intelligenz auf Verbrecherjagd: Einsatz von Gesichtserkennungstechnologie zur Aufklärung der „Kapitolverbrechen“, VerfBlog, 2021/1/22, abrufbar unter https://verfassungsblog.de/ki-verbrecherjagd/ (letzter Zugriff: 30.01.2022).

[Rü17] Rückert, C.: Zwischen Online-Streife und Online-(Raster-)Fahndung – Ein Beitrag zur Verarbeitung öffentlich zugänglicher Daten im Ermittlungsverfahren, ZStW 2017, 129 (2); 302–333.

[Sc15] Scherschel, F. A.: Darknet-Suchmaschine Memex: DARPA legt Projekt-Bausteine als Open Source offen, Heise Online (20.04.2015), abrufbar unter https://www.heise.de/security/meldung/Darknet-Suchmaschine-Memex-DARPA-legt-Projekt-Bausteine-als-Open-Source-offen-2613349.html (letzter Zugriff: 30.01.2022).

[Si17] Sierra, S.; Montes-y-Gómez, M.; Solorio, T.; González, F. A.: Convolutional Neural Networks for Author Profiling, Notebook for PAN at CLEF 2017, S. 1–7.

[SJ20] Staffler, L./Jany, O.: Künstliche Intelligenz und Strafrechtspflege – eine Orientierung, ZIS 2020, S. 164–177.

[St17] Stam, F.: Die Strafbarkeit des Aufbaus von Botnetzen, ZIS 2017, 547– 552.

[TP21] Taeger, J.; Pohle, J. (Hrsg.): Computerrechts-Handbuch. Informationstechnologie in der Rechts und Wirtschaftspraxis, 36. EL Februar 2021, C. H. Beck Verlag München.

[Ta13] Tanriverdi, H.: Mutmaßlicher Betreiber der Drogen-Plattform „Silk Road“: Drogenhandel, Mordversuch – und den Klarnamen angegeben, Süddeutsche Zeitung (04.10.2013), abrufbar unter https://www.sueddeutsche.de/digital/mutmasslicher-

betreiber-der-drogen-plattform-silk-road-drogenhandel-mordversuch-und-den-klarnamen-angegeben-1.1786870 (letzter Zugriff: 30.01.2022).

[Pa22] Tarnkappe.info: Palantir: Bayerisches Landeskriminalamt nutzt neue Analysesoftware (08.03.2022), abrufbar unter https://tarnkappe.info/artikel/datenschutz/palantir-bayerisches-landeskriminalamt-nutzt-neue-analysesoftware-216846.html (letzter Zugriff: 29.04.2022).

[Tz19] Tzanetakis, M.: Zum Phänomen der Drogenmärkte im Darknet, in: Tzanetakis/Stöver (Hrsg.), Drogen, Darknet und Organisierte Kriminalität. Herausforderungen für Politik, Justiz und Drogenhilfe, S. 11–21, 1. Auflage 2019, Nomos Verlag Baden-Baden.

[Va20] Van Note, C.: How OSINT Took Down the Dark Web's Silk Road, Scopenow (22.07.2020), abrufbar unter https://www.skopenow.com/news/how-osint-took-down-the-dark-webs-silk-road (letzter Zugriff: 30.01.2022).

[Wa12] Watson, K. D.: The Tor Network: A Global Inquiry into the Legal Status of Anonymity Networks, Washington University Global Studies Law Review 2012, Vol. 11, Issue 3, S. 715–737, abrufbar unter https://openscholarship.wustl.edu/law_globalstudies/vol11/iss3/6 (letzter Zugriff: 30.01.2022).

[Wi22] Wittmer, S.; Platzer, F., Steinebach M.; Yannikos, Y.: Deanonymisierung im Tor-Netzwerk – Technische Möglichkeiten und rechtliche Rahmenbedingungen (S. 151–168), in: Friedewald/Kreutzer/Hansen (Hrsg.), Selbstbestimmung, Privatheit und Datenschutz. Gestaltungsoptionen für einen europäischen Weg, Springer Vieweg Wiesbaden 2022.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 589

ML-basierte Klassifizierung von E-Mails für die datenschutzkonforme Löschung und Archivierung

Thomas Kunz, Ulrich Waldmann[1]

Abstract: E-Mails enthalten in der Regel personenbezogene Daten, die den datenschutzrechtlichen Löschvorgaben unterliegen. Eine angemessene Umsetzung der Löschvorgaben stellt jedoch die verantwortlichen Unternehmen vor eine große Herausforderung, zumal nach Erfüllung des Verarbeitungszwecks oftmals unterschiedliche (spezial-)rechtliche Aufbewahrungspflichten einer sofortigen Löschung entgegenstehen. Für die Einhaltung von Lösch- und Aufbewahrungspflichten ist es zunächst erforderlich, E-Mails, die diesen Verpflichtungen unterliegen (z. B. Rechnungen) zu identifizieren. Dieser Beitrag untersucht, inwieweit E-Mails mithilfe von maschinellem Lernen (ML) klassifiziert werden können. Für auf diese Weise klassifizierte E-Mails kann im nächsten Schritt entschieden werden, ob sie gemäß den Anforderungen der Datenschutz-Grundverordnung (DSGVO) gelöscht oder gemäß gesetzlicher Aufbewahrungsfristen länger aufbewahrt und archiviert werden müssen. Der Beitrag beschreibt zudem die Entwicklung eines Proof-of-Concept in Form eines Add-ons für Microsoft Outlook, das Nutzern erlaubt, die in ihren Postfächern enthaltenen E-Mails zu klassifizieren.[2]

Keywords: DSGVO; Löschpflicht; Aufbewahrungsfrist; E-Mail; Machine Learning; Klassifikation; AdaBoost; Naive Bayes; Random Forest

1 Einleitung

Gemäß Datenschutz-Grundverordnung (DSGVO) müssen Unternehmen, die personenbezogene Daten verarbeiten, nach Erreichung des Verarbeitungszwecks die betreffenden Daten löschen. Dieser Löschpflicht stehen bei bestimmten Daten (z. B. Gesundheitsdaten, Rechnungen) gesetzliche Aufbewahrungspflichten entgegen. Eine angemessene Umsetzung der Anforderungen stellt die Unternehmen insbesondere bei E-Mails vor eine große Herausforderung. Denn E-Mails sind oft Bestandteil wichtiger datenverarbeitender Prozesse, unterliegen unterschiedlichen Zwecken, werden aber oft nicht sauber nach Datentypen und Verarbeitungszwecken getrennt verwaltet. Lösch- und Aufbewahrungsfristen sind dann schwierig einzuhalten. Mögliche Lösungen sind das individuelle Sichten und Löschen jeder einzelnen E-Mail durch die Mitarbeitenden oder das pauschale Löschen nach der jeweils längsten einschlägigen Aufbewahrungsfrist. Die erste Lösung gilt als unangemessen

[1] Beide: Fraunhofer Institut für Sichere Informationstechnologie SIT | ATHENE, Rheinstr. 75, 64295 Darmstadt, Germany, thomas.kunz@sit.fraunhofer.de, ulrich.waldmann@sit.fraunhofer.de

[2] Diese Forschungsarbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF) und vom Hessischen Ministerium für Wissenschaft und Kunst (HMWK) im Rahmen ihrer gemeinsamen Förderung für das Nationale Forschungszentrum für angewandte Cybersicherheit ATHENE unterstützt.

doi:10.18420/inf2022_48

aufwändig, die zweite als nicht ausreichend zum Schutz der Betroffenen. Empfohlen wird stattdessen das automatische Verwalten individueller, thematisch gefüllter E-Mail-Ordner für jeden Posteingang. Dazu werden konfigurierte Löschregeln angewendet bzw. die Ordner zugriffsgeschützt archiviert, wobei nach Ablauf der konfigurierten Aufbewahrungsfrist der Postfachinhaber zur manuellen Überprüfung und Löschung aufgefordert wird. Dies setzt aber voraus, dass die Postfachinhaber ihre ein- und ausgehenden E-Mails regelmäßig in die korrekten Ordner verschieben und für diese Ordner auch die korrekten Löschzeitpunkte einstellen. Für eine unternehmensweit einheitliche Verwaltung von E-Mails müssen daher die Mitarbeitenden viel Disziplin aufbringen.[DSP19]

Allerdings haben Mitarbeitende das Wissen über Löschfristen und Aufbewahrungspflichten häufig gar nicht und sollten darüber hinaus nicht dazu verpflichtet sein, manuell E-Mails zu löschen bzw. zu archivieren. In der Praxis ordnen Mitarbeitende ihre E-Mails nicht einheitlich (wenn überhaupt), sondern haben dafür unterschiedliche Strategien. So würden Mitarbeitende dieselbe E-Mail unterschiedlich in verschiedene Ordner verschieben, beispielsweise dieselbe Bewerbungs-Mail eines Studenten für ein bestimmtes Projekt in den Ordner „Bewerbungen“ oder „Projekt xy“ oder „Studenten“, womit das korrekte Setzen von Ordner-spezifischen Löschregeln schwierig wird. Unser Lösungsansatz zielt daher darauf ab, E-Mails unabhängig von einer möglichen manuellen Sortierung durch ML-basierte Verfahren automatisch zu klassifizieren, um sensible Inhalte zu identifizieren und Mitarbeiter automatisch auf anstehende Lösch- und Archivierungsfristen hinzuweisen.

2 Datenschutzrechtliche Anforderungen

Der Grundsatz der Speicherbegrenzung gemäß Art. 5(1)(e) DSGVO besagt, dass personenbezogene Daten nur so lange, wie es für die Erreichung der Zwecke, für die sie erhoben wurden, erforderlich ist, in einer Form gespeichert werden dürfen, die eine Identifizierung der betroffenen Person ermöglicht. Sobald personenbezogene Daten für den Zweck, für den sie erhoben wurden, nicht mehr erforderlich sind, dürfen sie nur weiterverarbeitet werden, wenn sie für im öffentlichen Interesse liegende Archivierungszwecke, für wissenschaftliche oder historische Forschungszwecke oder für statistische Zwecke erforderlich sind. Wenn keine der Ausnahmen zutrifft, müssen die Daten gelöscht oder anonymisiert werden. Dieser Grundsatz wird in Art. 17 DSGVO durch das so genannte *Recht auf Löschung* konkretisiert. Diese Anforderungen gelten auch für personenbezogene E-Mails. Nach Erreichen des Verarbeitungszwecks müssen E-Mails grundsätzlich gelöscht oder für einen der oben genannten Zwecke mit eingeschränkten Zugriffsrechten archiviert werden. Zu beachten ist, dass bei E-Mails im Unternehmensumfeld der Arbeitgeber der Verantwortliche der Datenverarbeitung ist. Die aus Sicht des Datenschutzes betroffenen Personen sind die Sender und Empfänger der E-Mails, insbesondere aber auch diejenigen Personen, auf die sich die personenbezogenen E-Mail-Inhalte beziehen (z. B. im Falle einer weitergeleiteten Bewerbung).

In europäischen oder nationalen Gesetzen kann geregelt sein, dass bestimmte Daten für einen bestimmten Zeitraum aufbewahrt werden müssen, der über die in der DSGVO normierte Speicherbegrenzung hinausgehen kann. Diese so genannten *Aufbewahrungsfristen* sind beispielsweise in Steuergesetzen zu finden, wo die Aufbewahrungsfristen für Buchhaltungsunterlagen und ausgestellte Rechnungen leicht ein Jahrzehnt betragen können. Aufbewahrungsfristen sind mit der DSGVO vereinbar und werden sogar in Art. 17(3) DSGVO behandelt. Bevor personenbezogene Daten gelöscht werden, muss daher geprüft werden, ob für sie rechtsverbindliche Aufbewahrungsfristen bestehen. Wenn sich Aufbewahrungsfristen auf personenbezogene Daten auswirken, dürfen sie nicht länger gespeichert werden, als es die längste einschlägige Aufbewahrungsfrist erfordert.[ESS19]

Die Nutzung von ML-basierten Verfahren für eine automatisierte Durchsetzung von Lösch- und Aufbewahrungsfristen benötigt eine geeignete Rechtsgrundlage der Datenverarbeitung, wenn personenbezogene Trainingsdaten zum Einsatz kommen. Denn eine solche Wiederverwendung von Daten ist mit einer Änderung des ursprünglichen Zwecks der Datenverarbeitung verbunden und eine freiwillige Einwilligung der Betroffenen ist meist nicht mehr möglich. Eine Möglichkeit besteht in der Nutzung synthetischer Trainingsdaten, die den Echtdaten inhaltlich möglichst nahekommen. Zur Stärkung des Datenschutzes kann auch das so genannte *Federated Learning* dienen. Dabei werden die ML-Modelle lokal, also dort, wo die Rohdaten vorliegen, trainiert und dann zentral zusammengefasst. Wenn die zentrale Modellbildung die lokalen Modelle nicht personenbezogen auswerten kann (sondern die personenbezogene Auswertung nur lokal beim Nutzer geschieht), dann wären gemäß Art. 6(1)(f) DSGVO die Interessen der Betroffenen weniger beeinträchtigt und die Interessen des Verantwortlichen könnten überwiegen.[KH19]

Bei der Umsetzung eines ML-basierten Verfahrens zur Durchsetzung von Lösch- und Aufbewahrungspflichten handelt es sich um eine technische Maßnahme i. S. d. Art. 24(1) und Art. 32(1) DSGVO. Derartige Maßnahmen sind regelmäßig *angemessen* umzusetzen. In die Entscheidung über die Angemessenheit einer solchen Maßnahme sind u. a. die Risiken für die Betroffenen, die Implementierungskosten und der Stand der Technik zu berücksichtigen.

3 Stand der Technik

Überwachte ML-Verfahren werden in einer Vielzahl von Anwendungsszenarien eingesetzt, um komplexere Klassifizierungsprobleme zu lösen, z. B. bei der Filterung von Spam-E-Mails [SB18]. *Natural Language Processing* lässt sich mittels *Deep Neural Networks* auf E-Mails anwenden, um Phishing-Mails automatisch zu erkennen. Zum Trainieren kommen *Federated Learning* und das Modell *Long Short-Term Memory* zum Einsatz [SCO21]. Die Autoren von [BP14] wendeten Spam-Filter-Technologien auf die Klassifizierung von personenbezogenen Daten in eingehenden E-Mails an, um E-Mails regelbasiert zu verarbeiten und zu schützen. Sie erreichten mit *Support Vector Machines* die besten Ergebnisse. Andere Autoren entwickelten Lösungen, indem sie *Named Entity Recognition* mittels *Naive Bayes*

und *Convolutional Neural Networks* anwendeten, um personenbezogene Daten aus sozialen Medien und dem Internet zu erkennen und diese anschließend zu pseudonymisieren oder zu anonymisieren [DT18]. Eine ML-basierte Anwendung von *Named Entity Recognition* auf E-Mails, d. h. das Auffinden von personenbezogenen Daten in Form von Namen mit anschließender Anonymisierung der E-Mails, scheint allerdings wenig geeignet, um die Einhaltung der DSGVO (oder bestehender Aufbewahrungsfristen) zu gewährleisten [OT18].

Ein weiterer Lösungsansatz kombiniert unüberwachte und überwachte ML-Verfahren, um eingehende E-Mails anhand ihrer Inhalte halbautomatisch zu sortieren. Dazu werden die E-Mails mittels eines *K-Means*-Modells gruppiert und mittels überwachter ML-Verfahren (*Support Vector Machine*, *Logistic Regression*, *KNN Random Forests*, *Gradient Boosting*) klassifiziert. Die Lösung schlägt dem Nutzenden eine Ordnerstruktur vor. Der Nutzende kann diese bearbeiten, Ordner hinzuzufügen oder löschen und E-Mails zwischen Ordnern verschieben. Die Lösung passt die anfängliche Gruppierung an diese manuellen Änderungen an und klassifiziert dann die eingehenden Mails entsprechend automatisch.[EK21] Andere Autoren setzen verschiedene ML-Klassifizierungsalgorithmen wie *Support Vector Machines*, *Naive Bayes*, *Neural Networks* und *J48* für die Klassifizierung von E-Mails ein und empfehlen den Einsatz von *Boosting*-Verfahren, weil diese die Lernmerkmale von schwachen Klassifizierern durch die Merkmale von starken Klassifizierern ersetzen und damit bessere Ergebnissen ermöglichen [ARB21].

4 Klassifikation von E-Mails

Mit Hilfe von maschinellem Lernen wollen wir unterschiedliche Klassen von E-Mails erkennen, die häufig im Unternehmenskontext vorkommen und für die unterschiedliche Aufbewahrungs- bzw. Löschfristen gelten. Wir haben uns exemplarisch für die Klassen „Bewerbung“, „Krankmeldung“, „Rechnung“ sowie „Sonstige“ (für E-Mails, die nicht in die drei ersten Klassen fallen) entschieden, da diese in Unternehmen sehr verbreitet sind. Bei Bewerbungen handelt es sich im Vergleich zu den übrigen Klassen um längere und heterogene Texte, da hierbei auch Anhänge wie Bewerbungsschreiben, Lebensläufe usw. berücksichtigt werden. Zu dieser Klasse zählt auch die Korrespondenz im Rahmen einer Bewerbung (z. B. Nachfragen, Zusagen, Absagen, Einladungen). Krankmeldungen sind im Vergleich dazu in der Regel eher kurz und knapp formulierte Texte. Berücksichtigt werden in diesem Zusammenhang auch Nachfragen zu Krankmeldungen, Gesundmeldungen, Arbeitsunfähigkeitsbescheinigungen usw. Rechnungen sind dagegen eher strukturiert aufgebaut und haben einen sehr begrenzten Wortschatz mit Begriffen wie z. B. „Rechnung“, „Lieferung“, „Leistung“ oder „Umsatzsteuer“.

Abb. 1 zeigt die typischen Verarbeitungsschritte zur Erstellung eines Modells für maschinelles Lernen. Während der *Vorverarbeitung* werden die Rohdaten, also in unserem Fall die E-Mails, in ein einheitliches, von ML-Algorithmen verarbeitbares Format gebracht, u. a. durch Umwandlung in Kleinbuchstaben und Entfernen von Satzzeichen und so genannten Stoppwörtern, die für eine Klassifizierung irrelevant sind. Der *Vektorisierer* wandelt

die Daten in eine mathematische Darstellung. Während der *Klassifikation* wird mithilfe eines Klassifikators das Modell erstellt und angelernt. Abschließend wird im Rahmen der *Evaluation* das Modell getestet.

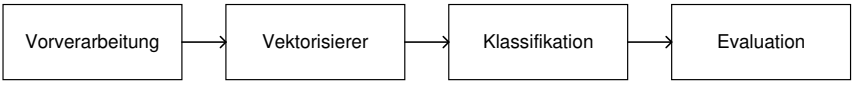


Abb. 1: Verarbeitungs-Pipeline

Der Schwerpunkt dieser Arbeit liegt auf der Klassifikation und Evaluation. Wir haben verschiedene Klassifikatoren auf ihre Eignung zur Identifizierung der oben genannten E-Mail-Klassen evaluiert. Zum Einsatz kamen Trainingsdaten im Umfang von insgesamt ca. 1.000 E-Mails aus den genannten Klassen. Es handelte sich hierbei sowohl um zuvor anonymisierte Echtdaten als auch um synthetische Daten. Die Klassifikatoren wurden mit Hilfe der 10-fachen Kreuzvalidierung evaluiert. Hierbei wird das Trainingsset in zehn gleich große Mengen aufgeteilt, dann mit neun Mengen davon trainiert und auf der übrig gebliebenen Menge getestet. Dies wird zehn Mal wiederholt, bis alle Mengen einmal getestet wurden.

Vor der eigentlichen Klassifikation findet die Vorverarbeitung der Daten statt. In diesem Rahmen werden die E-Mails zunächst in einzelne Wörter (Token) zerlegt. Von diesen Wörtern werden im zweiten Schritt Stoppwörter entfernt, d. h. Wörter, welche für die Klassifikation keine Bedeutung haben (z. B. „der“, „eine“, „und“, „aber“). Der letzte Schritt bildet das so genannte *Stemming*, d. h. das Zurückführen der Wörter auf ihren Wortstamm.

Im Rahmen der Klassifikation haben wir drei Klassifikatoren untersucht: *Naiver Bayes*, *Random Forest* und *AdaBoost*. Im Folgenden werden die Ergebnisse der Evaluation erläutert. Jeder Klassifikator wird anhand der Metriken *True Positive Rate (TP-Rate)*, *False Positive Rate (FP-Rate)*, *Precision*, *Recall*, *F-Measure* und *ROC* bewertet. Grundlage für die Metriken bilden die Anzahl der korrekt positiv klassifizierten Daten (*True Positives, TP*), die Anzahl der korrekt negativ klassifizierten Daten (*True Negatives, TN*), die Anzahl der falsch positiv klassifizierten Daten (*False Positives, FP*) und die Anzahl der falsch negativ klassifizierten Daten (*False Negatives, FN*).

TP-Rate (auch *Recall* genannt) bezeichnet den Anteil der vorhergesagten Positiven, die korrekt als echte Positive erkannt werden, von der Gesamtzahl der vorhergesagten Positiven: $\textit{TP-Rate} = \frac{TP}{TP+FN}$

FP-Rate bezeichnet den Anteil der fälschlicherweise als Positive erkannten von der Gesamtzahl aller tatsächlich Negativen: $\textit{FP-Rate} = \frac{FP}{FP+TN}$

Precision bezeichnet den Anteil der vorhergesagten Positiven, die korrekt als echte Positive erkannt werden, von der Gesamtzahl der tatsächlich Positiven: $\textit{Precision} = \frac{TP}{TP+FP}$

F-Measure ist ein Maß für die Genauigkeit des Tests und berechnet sich wiederum aus *Precision* und *TP-Rate*: $\textit{F-Measure} = 2 \cdot \frac{\textit{Precision} \cdot \textit{TP-Rate}}{\textit{Precision}+\textit{TP-Rate}}$

Die *ROC-Kurve* (Grenzwertoptimierungskurve) trägt in einem Diagramm die *TP Rate* (Y-Achse) gegen die *FP Rate* (X-Achse) anhand des variierten Grenzwerts für die jeweilige Klassenzuordnung auf und dient zur Beurteilung eines Klassifikators. ROC-Kurven im Bereich der unteren rechten Ecke weisen auf ein schlechtes Leistungsniveau des Klassifikators hin, während ROC-Kurven im Bereich der oberen linken Ecke auf eine gute Leistung hinweisen. Geht also der Wert der Fläche unter der Kurve gegen 1, d. h. die Kurve besteht annähernd aus zwei Geraden – vom Ursprung zur linken oberen Ecke und weiter zur rechten oberen Ecke, dann handelt es sich um einen sehr guten Klassifikator.

Naiver Bayes-Klassifikator. Der Naive Bayes-Klassifikator zählt zu den wahrscheinlichkeitsbasierten Klassifikatoren. Wir haben ihn ausgewählt, weil er sehr häufig bei der Textklassifikation, insbesondere bei der Erkennung von Spam-E-Mails eingesetzt wird.

Bei dem Test des Klassifikators mithilfe der Kreuzvalidierung wurden 87,33% der E-Mails korrekt klassifiziert (durchschnittliche TP-Rate). Die Konfusionsmatrix (Tabelle 1) sowie Tabelle 2 zeigen das Ergebnis des Tests anhand der eingangs erläuterten Metriken an. Es zeigt sich, dass der Naive Bayes-Klassifikator sehr gut in der Lage ist, Rechnungen korrekt zu klassifizieren, während er bei Krankmeldungen weniger gut abschneidet, siehe Tabelle 1, Spalte „Precision“. Dies bestätigen auch die entsprechenden ROC-Kurven (Abb. 2).

	Bewerbung (vorherges.)	**Krankmeldung** (vorhergesagt)	**Rechnung** (vorherges.)	**Sonstige** (vorherges.)
Bewerbung (tatsächlich)	196	3	0	13
Krankmeldung (tatsächlich)	3	127	0	27
Rechnung (tatsächlich)	0	0	357	4
Sonstige (tatsächlich)	23	60	5	271

Tab. 1: Konfusionsmatrix für den Naiven Bayes-Klassifikator

Klasse	**TP Rate**	**FP Rate**	**Precision**	**F-Measure**	**ROC**
Bewerbung	0,925	0,030	0,883	0,903	0,974
Krankmeldung	0,809	0,068	0,668	0,732	0,961
Rechnung	0,989	0,007	0,986	0,989	0,994
Sonstige	0,755	0,060	0,860	0,804	0,948
Durchschnitt	0,873	0,038	0,879	0,874	0,970

Tab. 2: Genauigkeit des Naiven Bayes-Klassifikators

Random Forest. Random Forest ist ein Vertreter der Klassifikatoren, die auf Entscheidungsbäumen basieren. Bei dem Test des Klassifikators mithilfe der Kreuzvalidierung

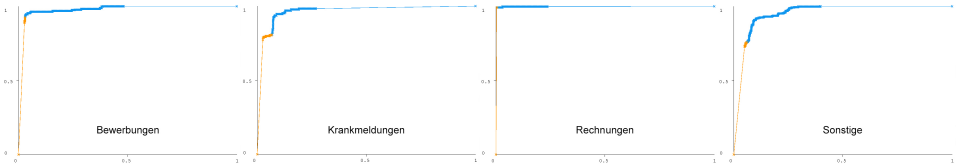


Abb. 2: ROC-Kurven (Naiver Bayes-Klassifikator)

wurden 97,52% der E-Mails korrekt klassifiziert (durchschnittliche TP-Rate). Die Konfusionsmatrix (Tabelle 3) sowie Tabelle 4 zeigen das Ergebnis des Tests anhand der oben genannten Metriken an. Es zeigt sich, dass der Random-Forest-Klassifikator sehr gut in der Lage ist, alle von uns gewählten E-Mail-Klassen korrekt zu klassifizieren, vgl. ROC-Kurven in Abb. 3.

	Bewerbung (vorherges.)	**Krankmeldung** (vorhergesagt)	**Rechnung** (vorherges.)	**Sonstige** (vorherges.)
Bewerbung (tatsächlich)	205	0	0	7
Krankmeldung (tatsächlich)	0	143	0	14
Rechnung (tatsächlich)	0	0	361	0
Sonstige (tatsächlich)	0	1	5	353

Tab. 3: Konfusionsmatrix für den Random-Forest-Klassifikator

Klasse	**TP Rate**	**FP Rate**	**Precision**	**F-Measure**	**ROC**
Bewerbung	0,967	0,000	1,000	0,983	0,999
Krankmeldung	0,911	0,001	0,993	0,950	0,998
Rechnung	1,000	0,007	0,986	0,993	1,000
Sonstige	0,983	0,029	0,944	0,963	0,994
Durchschnitt	0,975	0,012	0,976	0,975	0,998

Tab. 4: Genauigkeit des Random-Forest-Klassifikators

AdaBoost. AdaBoost gehört zur Familie der Boosting-Algorithmen. Er wurde ausgewählt, da Boosting-Algorithmen allgemein sehr gute Ergebnisse erzielen. Boosting-Algorithmen nutzen mehrere schwache Klassifikatoren, bei AdaBoost sogenannte Basis-Schätzfunktionen, um einen starken Klassifikator zu bilden. Die einzigen Bedingungen, welche schwache Klassifikatoren erfüllen müssen, ist eine höhere Genauigkeit als der Wurf einer Münze (Genauigkeit muss demnach über 50% liegen), als auch die Ausgabe eines probabilistischen Ergebnisses.

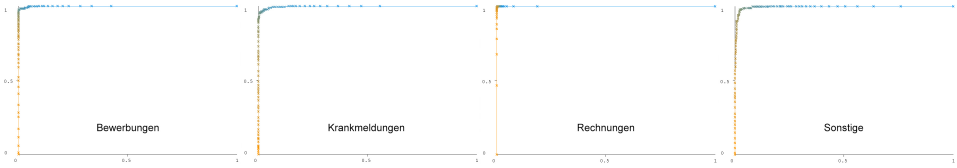


Abb. 3: ROC-Kurven (Random Forest)

In unserem Test konfigurierten wir AdaBoost mit dem auf Entscheidungsbäumen basierenden J48 als Basisklassifikator und evaluierten AdaBoost mit 10, 100 und 200 Iterationen. Es zeigte sich, dass in unserem Fall AdaBoost mit 100 Iterationen das beste Ergebnis lieferte und 98,2% der E-Mails korrekt klassifizierte (vgl. Konfusionsmatrix 5, Tabelle 6 sowie ROC-Kurven in Abb. 4). Bei 10 Iterationen wurden lediglich 96,79% der E-Mails korrekt klassifiziert. Eine Erhöhung der Iterationen auf 200 brachte dagegen keine zusätzliche Steigerung der korrekt klassifizierten E-Mails (98,07%). Insgesamt schneidet AdaBoost bei unseren Tests bei der Klassifikation von Bewerbungen, Krankmeldungen und Rechnungen besser ab als der Naive Bayes-Klassifikator. Bei Bewerbungen und Krankmeldungen schneidet AdaBoost zudem etwas besser ab als der Random-Forest-Klassifikator, dafür liefert Random Forest bei Rechnungen ein etwas besseres Ergebnis als AdaBoost.

	Bewerbung (vorherges.)	**Krankmeldung** (vorhergesagt)	**Rechnung** (vorherges.)	**Sonstige** (vorherges.)
Bewerbung (tatsächlich)	207	0	0	5
Krankmeldung (tatsächlich)	0	150	0	7
Rechnung (tatsächlich)	0	0	361	0
Sonstige (tatsächlich)	3	3	2	351

Tab. 5: Konfusionsmatrix für den AdaBoost-Klassifikator (Basisklassifikator: J48, Iterationen: 100)

Klasse	**TP Rate**	**FP Rate**	**Precision**	**F-Measure**	**ROC**
Bewerbung	0,976	0,003	0,986	0,981	0,993
Krankmeldung	0,955	0,003	0,980	0,968	0,997
Rechnung	1,000	0,003	0,994	0,997	0,999
Sonstige	0,978	0,016	0,967	0,972	0,993
Durchschnitt	0,982	0,007	0,982	0,982	0,996

Tab. 6: Genauigkeit des AdaBoost-Klassifikators (Basisklassifikator: J48, Iterationen: 100)

Fazit. Nach Abschluss der Evaluation der drei Klassifikatoren *Naiver Bayes-Klassifikator*, *Random Forest* und *AdaBoost* erweist sich AdaBoost mit J48 als Basisklassifikator und 100 Iterationen im Durchschnitt als der geeignetste Klassifikator für die Identifizierung

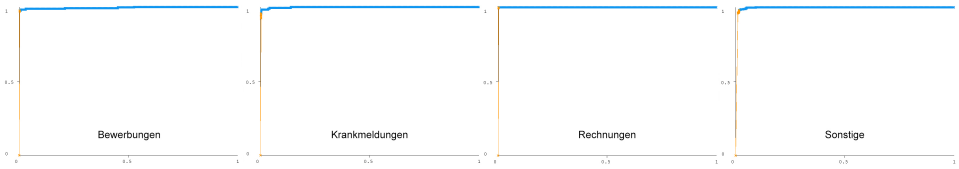


Abb. 4: ROC-Kurven (AdaBoost, J48, 100 Iterationen)

der von uns gewählten E-Mail-Klassen. Er erzielte bei fast allen von uns berücksichtigten Metriken etwas bessere durchschnittliche Kennzahlen als die anderen beiden Klassifikatoren. Zu beachten ist jedoch, dass er nur ein geringfügig besseres Ergebnis als Random Forest erreichte.

5 Weiterverarbeitung der klassifizierten E-Mails

Nachdem die E-Mails wie in Abschnitt 4 beschrieben klassifiziert wurden, muss im nächsten Schritt ermittelt werden, bei welchen E-Mails die gesetzlichen Löschfristen abgelaufen sind und die E-Mails daher zu löschen sind. Grundsätzlich ist davon auszugehen, dass E-Mails, die zu den von uns gewählten Klassen zählen, im Allgemeinen personenbezogene Daten enthalten. Demnach gilt gemäß Art. 5(1)(e) DSGVO, dass diese E-Mails nach Erreichen des ursprünglichen Verarbeitungszwecks unverzüglich zu löschen sind. Dem gegenüber gelten für diese E-Mail-Klassen jedoch zusätzlich spezielle Aufbewahrungspflichten, die zu berücksichtigen sind. Im Folgenden werden wir die üblichen Aufbewahrungsfristen für Rechnungen, Krankmeldungen und Bewerbungen kurz nennen. Zu beachten ist jedoch, dass es in einzelnen Organisationen oder Unternehmen aufgrund weiterer Vorgaben Abweichungen von den genannten Aufbewahrungsfristen geben kann.

Rechnungen sind gemäß § 147 AO und § 14b UStG für einen Zeitraum von zehn Jahren aufzubewahren, bevor sie gelöscht werden sollten. Als Berechnungsgrundlage für das Ende des Aufbewahrungszeitraums verwenden wir das Absendedatum der E-Mail. In Bezug auf Krankmeldungen per E-Mail und Arbeitsunfähigkeitsbescheinigungen gilt Folgendes: Diese gehören zu den Unterlagen, die bei einer Betriebsprüfung, die gemäß § 28p SGB IV mindestens alle vier Jahre stattfindet, vorgelegt werden müssen. Daher gehen wir im Falle von Krankmeldungen von einer Mindestaufbewahrungsfrist von vier Jahren aus. Im Falle von Bewerbungen wird empfohlen, diese spätestens sechs Monate nach Beendigung des Bewerbungsverfahrens zu löschen.[3] Diese Frist besteht, da der Bewerber gemäß Allgemeinem Gleichbehandlungsgesetz (AGG) die Möglichkeit haben soll, gegen eine Absage zu klagen. Eine Ausnahme von dieser Löschfrist besteht, wenn der Bewerber seine Einwilligung gegeben hat, dass seine Daten länger gespeichert werden

[3] Centralvereinigung Deutscher Wirtschaftsverbände für Handelsvermittlung und Vertrieb (CDH): „DSGVO: Aufbewahrungs- und Löschfristen für Bewerbungsunterlagen“, https://cdh.de/themenfeld/dsgvo-aufbewahrungs-und-loeschfristen-fuer-bewerbungsunterlagen/

dürfen, zum Beispiel weil der Bewerber möchte, dass das Unternehmen seine Bewerbung bei zukünftigen Stellenausschreibungen berücksichtigt. Bei Bewerbungen kann demnach nicht das Absendedatum verwendet werden, um zu beurteilen, ob eine E-Mail gelöscht werden muss. Stattdessen muss der Zeitpunkt der Beendigung des jeweiligen Bewerbungsverfahrens als Grundlage herangezogen werden und es muss zudem geprüft werden, ob der Bewerber seine Einwilligung für eine längerfristige Aufbewahrung gegeben hat.

E-Mails, deren Aufbewahrungsfrist abgelaufen ist, sollten unverzüglich gelöscht werden. Dem Nutzer sollte jedoch die Möglichkeit gegeben werden, die zu löschenden E-Mails zuvor anzuschauen und sich ggf. gegen eine Löschung zu entscheiden. Ist die Aufbewahrungsfrist dagegen noch nicht abgelaufen, dürfen die E-Mails nicht gelöscht werden. Für diese E-Mails schlagen wir vor, sie mit einem speziellen Label zu versehen, wie es verschiedene E-Mail-Clients wie z. B. Microsoft Outlook oder Mozilla Thunderbird bieten. Dieses Label enthält das ermittelte Löschdatum der jeweiligen E-Mail. Gelabelte E-Mails müssen zukünftig nicht erneut klassifiziert werden, stattdessen genügt es, automatisiert in regelmäßigen Abständen das in den Labels enthaltene Löschdatum zu prüfen. Ein automatisches Verschieben der E-Mails, deren Aufbewahrungsfrist noch nicht abgelaufen ist, in einen speziellen Ordner erscheint dagegen nicht vorteilhaft: Es ist in der Regel unbekannt, ob die Bearbeitung der E-Mails durch den Nutzer abgeschlossen ist und ein automatisches Verschieben würde vermutlich dazu führen, dass der Nutzer viele E-Mails nicht mehr wiederfindet.

6 Proof-of-Concept

Auf Grundlage der in Abschnitt 4 beschriebenen Evaluation der Klassifikatoren wurde im Anschluss ein Proof-of-Concept in Form eines Add-ons für Microsoft Outlook implementiert. Mithilfe dieses Add-ons kann das zuvor angelernte Modell anhand von Echtdaten getestet werden. Das Modell für das maschinelle Lernen wurde in Python unter Verwendung der Bibliotheken *scikit-learn* und *spaCy* implementiert. Das eigentliche Add-on wurde in C# entwickelt. Das Add-on bietet dem Nutzer die Möglichkeit, die in seinem E-Mail-Postfach enthaltenen E-Mails anhand der Klassen Bewerbung, Krankmeldung, Rechnung sowie Sonstige zu klassifizieren. Hierzu selektiert der Nutzer zunächst diejenigen Mailordner, die bei der Klassifizierung berücksichtigt werden sollen. Das Add-on übergibt daraufhin die in den Ordnern gespeicherten E-Mails an die in Python implementierte Machine-Learning-Komponente, welche die Klassifikation durchführt und das Ergebnis an das Add-on zurückliefert. Das Add-on zeigt dem Nutzer in einem separaten Fenster das Ergebnis der Klassifikation an (siehe Screenshot in Abb. 5). Hierbei wird auch das Alter einer E-Mail berücksichtigt, d. h. es wird geprüft, ob die Aufbewahrungsfrist abgelaufen ist. Solche E-Mails werden in rot dargestellt. Der Nutzer hat nun die Möglichkeit, die E-Mails zu öffnen und diejenigen E-Mails zu markieren, die er entweder direkt löschen möchte oder die er zunächst in einen speziellen Archivordner verschieben möchte.

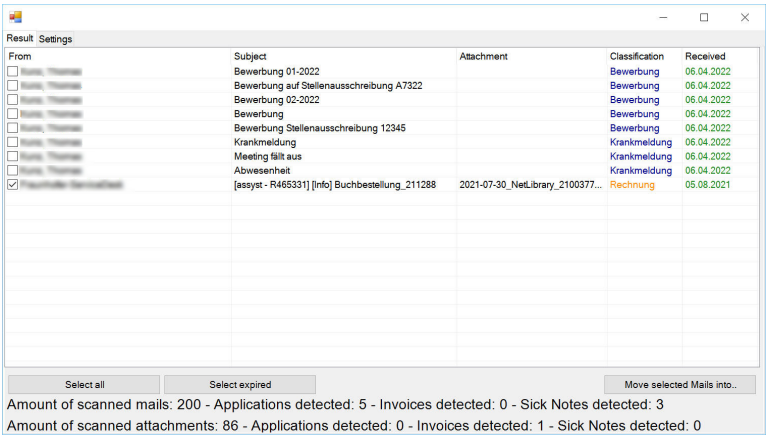


Abb. 5: Add-on: Beispiel eines Klassifikationsergebnisses

7 Zusammenfassung und Ausblick

Gemäß dem in der DSGVO festgelegten Grundsatz der Speicherbegrenzung müssen personenbezogene Daten gelöscht werden, wenn der eigentliche Verarbeitungszweck nicht mehr besteht. Allerdings stehen dieser Löschfrist häufig (spezial-)rechtliche Aufbewahrungspflichten gegenüber, die dafür sorgen, dass personenbezogene Daten auch nach Erreichen des datenschutzrechtlichen Verarbeitungszwecks aufbewahrt werden müssen. In diesem Beitrag haben wir uns mit der Frage beschäftigt, inwieweit typische E-Mails aus dem Unternehmenskontext mit Hilfe von maschinellem Lernen identifiziert und nach Ablauf ihrer (spezial-)gesetzlichen Aufbewahrungsfristen automatisiert gelöscht werden können. Zu diesem Zweck haben wir exemplarisch die E-Mail-Klassen „Bewerbung“, „Krankmeldung“ und „Rechnung“ ausgewählt und ein Modell für das maschinelle Lernen erstellt, das diese E-Mail-Klassen identifizieren soll. Es hat sich im Rahmen der Evaluation verschiedener Klassifikatoren gezeigt, dass insbesondere AdaBoost, aber auch der Random Forest Klassifikator, sehr gut in der Lage ist, zumindest diese Klassen korrekt zu erkennen.

In einem nächsten Schritt wollen wir die Evaluation der Klassifikatoren um weitere Unterlagen und Dokumente ausweiten, die sehr häufig in Form von E-Mails vorliegen. Beispiele hierfür sind Buchungsbelege, Mahnungen, Arbeitsverträge und Arbeitszeugnisse, für die jeweils unterschiedliche Aufbewahrungsfristen gelten. In diesem Zusammenhang sollen auch weitere Klassifikatoren (z. B. XGBoost) evaluiert werden, insbesondere falls sich herausstellen sollte, dass die bisher untersuchten Klassifikatoren bei anderen E-Mail-Klassen weniger gut abschneiden. Bislang wurden für das Anlernen des Modells hauptsächlich synthetische Daten verwendet, die in der Regel eine schlechtere Qualität haben als Echtdaten. Wir wollen daher untersuchen, inwieweit auch Echtdaten als Trainingsdaten für das Anlernen eines Modells genutzt werden können, ohne hierbei gegen Vorgaben der DSGVO zu verstoßen, z. B. hinsichtlich der Verarbeitung oder Weitergabe von personenbezogenen Daten. Ein

möglicher Ansatz hierfür stellt das so genannte *Federated Learning* dar. Hierbei werden lokal auf den Rechnern der einzelnen Nutzer jeweils Modelle unter Verwendung deren E-Mails trainiert. Diese (Teil-)Modelle werden anschließend zu einem gemeinsamen Modell zusammengefügt.

Die vorgestellte Lösung automatischer E-Mail-Klassifizierung könnte auch für die Verwaltung bestehender E-Mail-Archive und für andere zentrale E-Mail-Verarbeitungsprozesse optimiert werden, um den Lösch- und Aufbewahrungsfristen angemessen nachkommen zu können. Eine 100% korrekte Klassifizierung der E-Mails wird aber wahrscheinlich weder client- noch serverseitig erreicht werden. Dennoch stellt eine solche automatisierte Lösung eine gute Alternative zu den bisherigen manuell-basierten (und keinesfalls weniger fehlerbehafteten) Lösungen dar, zumal eine Ausfilterung besonders sensibler E-Mails das Risiko für die betroffenen Personen erheblich senken kann. Eine nahezu fehlerfreie automatisierte Lösung kann daher i. S. d. Art. 24(1) DSGVO als angemessen gelten.

Literatur

[ARB21] Anitha, P.; Rao, C. G.; Babu, D. S.: Email Spam Filtering Using Machine Learning Based Xgboost Classifier Method. Turkish Journal of Computer and Mathematics Education 12/11, S. 2182–2190, 2021.

[BP14] Bier, C.; Prior, J.: Detection and labeling of personal identifiable information in e-mails. In: IFIP SEC. Springer, S. 351–358, 2014.

[DSP19] Durmus, E.; Selzer, A.; Pordesch, U.: Das Löschen nach der DSGVO. Datenschutz und Datensicherheit-DuD 43/12, S. 786–791, 2019.

[DT18] Di Cerbo, F.; Trabelsi, S.: Towards personal data identification and anonymization using machine learning techniques. In: European Conference on Advances in Databases and Information Systems. Springer, S. 118–126, 2018.

[EK21] Erker, J.-J.; Koehler, M.: Combining Unsupervised and Supervised Machine Learning for Semi-automatic E-Mail Classification./, 2021.

[ESS19] Enzmann, M.; Selzer, A.; Spychalski, D.: Data Erasure under the GDPR-Steps towards Compliance. Eur. Data Prot. L. Rev. 5/, S. 416, 2019.

[KH19] Kaulartz, M.; Huth, M.: Datenschutz-Compliance bei Machine Learning. Datenschutz-Berater 12/, S. 276–278, 2019.

[OT18] Olby, L.; Thomander, I.: A Step Toward GDPR Compliance: Processing of Personal Data in Email, KTH Royal Institute of Technology, Stockholm, 2018.

[SB18] Sharma, P.; Bhardwaj, U.: Machine Learning based Spam E-Mail Detection. International Journal of Intelligent Engineering and Systems 11/3, 2018.

[SCO21] Sun, Y.; Chong, N.; Ochiai, H.: Privacy-Preserving Phishing Email Detection Based on Federated Learning and LSTM. preprint arXiv:2110.06025/, 2021.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 601

Technische und Rechtliche Auseinandersetzung mit Weiterleitungs-URLs in E-Mails aus Security-Gründen

Dirk Müllmann[1], Maxime Veit[2], Melanie Volkamer [3]

Abstract: Häufig sind Links in E-Mails eingebunden, um Empfänger*innen einfach auf Webseiten oder Webseiteninhalte hinweisen zu können. Dazu wird hinter einem Link die entsprechende URL (auch Webadresse genannt) hinterlegt. Zunehmend ist zu beobachten, dass es sich bei der hinterlegten URL jedoch nicht um die sog. Ziel-URL, d.h. die eigentliche Adresse der Webseite, handelt, sondern eine sog. Weiterleitungs-URL hinterlegt wurde. Anders als die Ziel-URL leitet die Weiterleitungs-URL zunächst auf eine andere URL weiter. Sie tritt in zwei unterschiedlichen Formen auf, die auch kombiniert werden können: Einerseits in Form von Weiterleitungs-URLs, die der Mailserver der Empfänger*innen aus Security-Gründen integriert, und anderseits in solcher, bei der die Weiterleitungs-URL von Absender*innen aus Marketinggründen verwenden werden. Ziel dieses Aufsatzes ist es, die Gruppe der Weiterleitungs-URLs aus Security-Gründen aus technischer und rechtlicher Sicht zu untersuchen und Empfehlungen für ihren Einsatz abzuleiten.

Keywords: URL-Security-Check, E-Mail, IT-Sicherheit, Datenschutz, Weiterleitungs-URL

1 Beschreibung

Die Möglichkeit, Links in E-Mails einzubetten, birgt den Vorteil der schnellen Auffindbarkeit konkreter Webseiten. Sie hat aber auch Nachteile: Angreifer*innen, häufig als Phisher*in bezeichnet, verschicken authentisch wirkende E-Mails, bei denen hinter den Links Phishing-URLs hinterlegt sind. Klicken Empfänger*innen auf einen solchen Link, führt die URL sie zu einer Phishing-Webseite. Diese versucht entweder Schadsoftware auf das Gerät der Empfänger*innen zu laden oder versucht möglichst authentisch auszusehen, damit Empfänger*innen dort sensible Daten, wie z.B. die Login-Daten der Originalseite eingeben.

Diese Form des Cyber-Angriffs stellt eine zunehmende Gefahr für Unternehmen und Privatpersonen dar.[4] Dies liegt unter anderem daran, dass die E-Mails immer authentischer von den Angreifer*innen gestaltet werden und Phishing-Mails häufig nur noch an den URLs hinter den Links als solchen entlarvt werden können. Die URL wird an Desktop-

[1] Dirk Müllmann ist Wissenschaftlicher Mitarbeiter am Institut für die Entwicklung sicherer Systeme am KIT sowie an der Goethe-Universität Frankfurt/Main am Lehrstuhl von Prof. Spiecker gen. Döhmann.
[2] Maxime Veit ist Wissenschaftlicher Mitarbeiter am KIT bei der Forschungsgruppe SECUSO.
[3] Prof. Dr. Melanie Volkamer ist ordentliche Professorin am Karlruher Institut für Technologie (KIT) und Leiterin der Forschungsgruppe SECUSO.
[4] Wirtschaftsschutz 2021. (2021). bitkom. https://www.bitkom.org/sites/default/files/2021-08/bitkom-slides-wirtschaftsschutz-cybercrime-05-08-2021.pdf.

doi:10.18420/inf2022_49

Geräten und Laptops in der sog. Statusleiste und je nach E-Mail-Client zusätzlich in einem Tooltip angezeigt (siehe Abbildung 1). Auf mobilen Geräten ist es in der Regel möglich, durch ein längeres Berühren des Links an die Information zur URL zu gelangen.

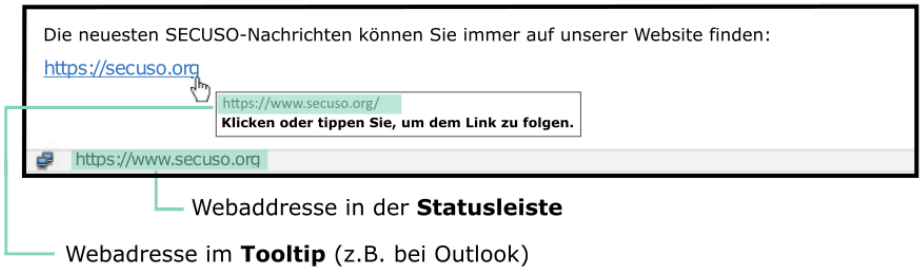


Abb. 1: Linkziel im Tooltip und Statusleiste

Eingehende E-Mails werden zwar üblicherweise beim Ankommen am Mailserver einem Security-Check unterzogen, der auch in der E-Mail enthaltene URLs prüft. Häufig sind die URLs zu diesem Zeitpunkt aber noch nicht als Phishing-URL bekannt. Somit kann die Mail nicht automatisch als Phishing-Mail erkannt und entsprechend behandelt werden. Als Konsequenz landen diese Phishing E-Mails in den Posteingängen der Empfänger*innen.

Um das Risiko, Opfer von Phisher*innen zu werden, zu minimieren, werden sowohl in der Praxis als auch in der Wissenschaft bisher zwei Ansätze verfolgt: (1) Die Einführung von Security-Awareness-Maßnahmen, mit denen die Teilnehmenden lernen, Phishing-Mails – insbesondere anhand der URL hinter den Links – zu erkennen[5] und (2) die Einführung von Security-Indikatoren, die Empfänger*innen helfen die URL vor dem Klicken des Links zu prüfen.[6] Zunehmend setzen Unternehmen und Organisationen auch einen dritten Ansatz ein: (3) Die Ziel-URLs werden durch entsprechende Security-Weiterleitungs-URLs ersetzt. Dies kann entweder vor oder nach Zustellung der E-Mail stattfinden. Entweder wird der Security-Check so erweitert, dass der Mailserver eine E-Mail nach ihrem Eingang zunächst wie bisher überprüft, im Fall der Zustellung der E-Mail aber alle Ziel-URLs durch entsprechende Weiterleitungs-URLs ersetzt, oder die E-Mail

[5] Vgl. Reinheimer, B. M., Aldag, L., Mayer, P., Mossano, M., Düzgün, R., Lofthouse, B., Von Landesberger, T., & Volkamer, M. (2020). An investigation of phishing awareness and education over time: When and how to best remind users [PDF]. https://doi.org/10.5445/IR/1000122566 ; Vgl. Tschakert, K. F., & Ngamsuriyaroj, S. (2019). Effectiveness of and user preferences for security awareness training methodologies. Heliyon, 5(6), e02010. https://doi.org/10.1016/j.heliyon.2019.e02010.

[6] Vgl. Volkamer, M., Renaud, K., Reinheimer, B., & Kunz, A. (2017). User experiences of TORPEDO: TOoltip-poweRed Phishing Email DetectiOn. Computers & Security, 71, 100–113. https://doi.org/10.1016/j.cose.2017.02.004 ; Vgl. Petelka, J., Zou, Y., & Schaub, F. (2019). ; Put Your Warning Where Your Link Is: Improving and Evaluating Email Phishing Warnings. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–15. https://doi.org/10.1145/3290605.3300748.

wird zugestellt und die Ersetzung der Ziel-URLs erfolgt erst bei der Anzeige der E-Mail im E-Mail-Client oder im Webbrowser.

Gegenstand der vorliegenden Untersuchung ist der dritte Ansatz, das Ersetzen der Ziel-URLs hinter Links durch Security-Weiterleitungs-URLs. Hierbei wird nach dem Klick auf den Link und vor der Weiterleitung an die Ziel-URL diese zunächst durch einen URL-Security-Check geprüft und nur dann aufgerufen, wenn die Ziel-URL als niedriges Risiko eingestuft wird. Grundsätzlich kann der Security-Check von dem Anbieter, der auch den Mailserver betreibt, oder von einem externen Anbieter durchgeführt werden.



Abb. 2: Weiterleitungs-URL in Statusleiste und Ziel-URL im Tooltip

Um vor dem Aufruf der Ziel-URL den URL-Security-Check durchführen zu können, wird im vorderen Teil der Security-Weiterleitungs-URL (d.h. insbesondere der Bereich der Second-Level-Domain und der Top-Level-Domain) auf den Server, der diesen URL-Security-Check durchführt, verwiesen. Dieser Teil ist somit in jeder Security-Weiterleitungs-URL identisch. Der hintere Teil der URL, d.h. die Information im Pfad oder Query, ist jeweils unterschiedlich. Je nach eingesetztem URL-Security-Check kann die Ziel-URL im hinteren Teil der URL noch enthalten sein. Falls nicht, sieht der hintere Teil der URLs wie eine zufällige Zeichenfolge aus. Je nach Umsetzungsform des E-Mail-Clients wird im Tooltip entweder die ursprüngliche Ziel-URL angezeigt und die Security-Weiterleitungs-URL ist nur in der Statusleiste zu sehen (siehe Abbildung 2) oder die Empfänger*innen haben nur Zugriff auf die Security-Weiterleitungs-URL hinter dem Link, nicht jedoch auf die Ziel-URL (siehe Abbildung 3).



Abb. 3: Weiterleitungs-URL in Statusleiste und im Tooltip

Durch die Verwendung des URL-Security-Checks wird die Wahrscheinlichkeit erhöht, Phishing-URLs zu entdecken und deren Aufruf zu verhindern. Der Vorteil gegenüber dem

Security-Check vor der Zustellung der E-Mail durch den Mailserver ist, dass in der Zwischenzeit die Ziel-URL als Phishing-URL bekannt geworden sein kann. Allerdings bietet auch dies keinen 100%igen Schutz. Insbesondere, wenn E-Mails unmittelbar nach dem Erhalt der Nutzer*innen bearbeitet werden, ist es fraglich, ob eine URL in der kurzen Zeitspanne als Phishing-URL gemeldet wird und der URL-Security-Check zu einem anderen Ergebnis kommt als der ursprüngliche Security-Check.

2 Technische Bewertung

In diesem Unterkapitel wird der Ansatz aus technischer Sicht bewertet, wobei der Fokus auf Security und Usability Problemen liegt.

2.1 Zusammenspiel mit Security-Awareness-Maßnahmen

Im Rahmen von Security-Awareness-Maßnahmen wird erklärt, dass viele Phishing-Mails an der Plausibilität von Absender*innen und Inhalt erkannt werden können und dass alle Phishing-Mails mittels Überprüfung der URL hinter dem Link entlarvt werden können. Für diese Überprüfung wird erläutert, dass es bei der Kontrolle der URL insbesondere auf die Prüfung der Kombination aus Second-Level-Domain und Top-Level-Domain (dem sog. Wer-Bereich) ankommt (siehe Abbildung 4).[7] Dieses Erkennungsmerkmal ist jedoch hinfällig, wenn die Ziel-URLs nicht angezeigt werden, weil sie durch Weiterleitungs-URLs ersetzt wurden, die immer aus der gleichen Kombination von Second-Level-Domain und Top-Level-Domain bestehen, da sie immer zum gleichen URL-Security-Check führen. Dies gilt je nach Umgebung für alle URLs bzw. für diejenigen, die in weitergeleiteten E-Mails sind.



Abb. 4: Kombination aus Second-Level-Domain und Top-Level-Domain

Die Awareness-Maßnahmen müssen also um Hinweise zur Funktionsweise des neuen URL-Security-Checks erweitert werden. Für den Fall, dass Adressat*innen der Security Maßnahmen nur E-Mails an einem E-Mail-Client lesen, der im Tooltip die ursprüngliche Ziel-URL und die Weiterleitungs-URL in der Statusbar anzeigt (siehe Abbildung 2), kann die Erklärung der Security-Awareness-Maßnahme zur Prüfung von URLs beibehalten werden, solange es sich nicht um weitergeleitete E-Mails handelt. Für diese Fälle sowie

[7] vgl. Betrügerische Nachrichten. (2020). SecUSo. https://secuso.aifb.kit.edu/downloads/Flyer/NoPhish_betr.Nachrichten/KIT-Faltblatt-BN-DE_08.11.2020.pdf.

für den Fall, dass Adressat*innen E-Mails auch in mobilen Umgebungen abrufen, in denen ihnen die Ziel-URL nicht angezeigt wird, müssten die Inhalte der Security-Awareness-Maßnahme angepasst werden.

Falls die Weiterleitungs-URL so gestaltet ist, dass die Ziel-URL im hinteren Teil der URL zu finden ist, sollte in der Security-Awareness-Maßnahme erklärt werden, wie die URL und insbesondere der Wer-Bereich im entsprechenden hinteren Teil der Weiterleitungs-URL geprüft werden kann. Zudem könnte das Problem auftreten, dass der hintere Teil der Weiterleitungs-URL (und damit der für die Prüfung relevante Aspekt) möglicherweise nicht vollständig in der Statusleiste zu sehen ist, da der Platz in der Statusleiste für die komplette URL nicht ausreicht. Auch hierauf ist in der Security-Awareness-Maßnahme hinzuweisen.

Falls die Weiterleitungs-URL hingegen keine Informationen über die Ziel-URL enthält und diese auch nicht angezeigt wird, ist die Erklärung zur Prüfung der URL vor dem Öffnen eines Links nicht mehr relevant und sollte aus der Security-Awareness-Maßnahme entfernt werden. In diesem Fall würden die Adressat*innen nur noch das Absender-Feld und den Inhalt der E-Mail auf Plausibilität prüfen und müssten sich im Übrigen auf den URL-Security-Check verlassen. Dieser kann die Ziel-URL anders als Empfänger*innen der E-Mail nicht im erweiterten Kontext auf Plausibilität prüfen, also etwa, ob der Link zur restlichen E-Mail passt oder ob Empfänger*innen die Webseite hinter der Ziel-URL kennen. Dies schränkt die Empfänger*innen von E-Mails in der Umsetzung eines effektiven Schutzes vor Phishing-Angriffen ein. Darüber hinaus kann es, sofern Security-Awareness-Maßnahmen bereits durchgeführt wurden, für die Anwender*innen verwirrend sein und sich negativ auf die Effektivität der Maßnahme auswirken, wenn erklärt wird, dass die zuvor erlernte eigene Überprüfung mit dem neuen Security-Check nicht mehr angewendet werden kann. Im schlimmsten Fall hat dies zur Folge, dass Empfänger*innen sich blind auf den URL-Security-Check verlassen und jeden Link unreflektiert anklicken. Die menschliche Firewall zur Prüfung der Links vor dem Klick fiele dann weg. Für diesen Fall gilt außerdem, dass ein sog. Mismatch zwischen einer im Linktext angegebenen URL und einer Ziel-URL (siehe Abbildung 5), anders als in der Security-Awareness-Maßnahme hingewiesen, nicht mehr zwangsläufig auf eine Phishing-URL hindeutet, sondern in der ständigen Ersetzung der Ziel-URL durch den URL-Security-Check begründet ist. Daher ist dieser Inhalt der Security-Awareness-Maßnahme zu aktualisieren – ggf. mit den bereits genannten negativen Konsequenzen.



Abb. 5: Link Mismatch zwischen Link-URL und Ziel-URL. Phisher*innen versuchen die Opfer durch eine vertrauenswürdige URL im Text in die Irre zu führen.

Insgesamt ist vor der Einführung von Security-Weiterleitungs-URLs also zunächst zu prüfen, inwieweit Security-Awareness-Maßnahmen gegenläufig sind und angepasst werden müssen.

Eine weitere große Herausforderung im Kontext der Einführung eines URL-Security-Checks besteht darin, dass bei Adressat*innen nicht der Eindruck entstehen sollte, dass durch die Einführung des URL-Security-Checks alle Links sicher angeklickt werden können, weil der URL-Security-Check die Prüfung übernimmt. Wenn es nicht gelingt, dies klar zu kommunizieren, würde das Security-Niveau durch die Einführung des URL-Security-Checks sogar reduziert und nicht gesteigert werden.

Hinzu kommt, dass Studien fehlen, die untersuchen, ob Nutzer*innen nach Security-Awareness-Maßnahmen Phishing-URLs besser erkennen als ein solcher URL-Security-Check. Es ist anzunehmen, dass die Ergebnisse solcher Studien stark vom Kontext, wie etwa unterschiedlicher Nutzergruppen und ob zusätzlich die Ziel-URL verfügbar ist oder nicht, abhängen und die Frage nur im Einzelfall beantwortet werden kann.

2.2 Probleme durch weitergeleitete bzw. beantworte E-Mails

Dadurch, dass der URL-Security-Check die URL hinter dem Link vor der Zustellung der E-Mail verändert, wird im Falle einer Weiterleitung bzw. Beantwortung der E-Mail die Weiterleitungs-URL zur Ziel-URL. Wenn diese E-Mails an Empfänger*innen geschickt werden, die nicht den gleichen URL-Security-Check nutzen, dann sorgt die Ziel-URL (ursprüngliche Weiterleitungs-URL) hinter den Links für Irritationen, da je nach Form des Links ein Link-Mismatch festgestellt wird und die Ziel-URL nicht wie gewohnt überprüft werden kann.

Das reduziert das Security-Niveau der Empfänger*innen. Wenn diese auf den Link klicken, wird die URL zunächst auch vom URL-Security-Check der Absender*innen geprüft. Sie können aber in Ermangelung von Informationen über das Angebot nicht einschätzen, inwieweit sie sich auf diesen URL-Security-Check verlassen können. Dieses Problem kann jedoch gelöst werden, indem die Weiterleitungs-URL beim Verschicken senderseitig wieder in die ursprüngliche Ziel-URL umgewandelt wird.

2.3 Probleme mit dem Einsatz von S/MIME und PGP

Einige Unternehmen und Privatpersonen nutzen zum Signieren und Verschlüsseln ihrer E-Mails S/MIME oder PGP. Dies erlaubt es Sender*innen E-Mails digital zu signieren, selbst wenn Empfänger*innen kein entsprechendes Schlüsselpaar besitzen. Wenn nun eine digital signierte E-Mail beim Mailserver ankommt und der Security-Check entscheidet, dass die E-Mail zugestellt wird, würde, sofern die Ersetzung nicht erst im E-Mail-Client realisiert wird, dieser auch hier die URLs ersetzen. Dies hat allerdings zur Folge, dass E-Mail-Clients eine Security-Intervention anzeigen, weil die digitale Signatur durch die nachträgliche Veränderung der E-Mail ungültig wird. Nun könnte zwar im Rahmen der

Security-Awareness-Maßnahme erklärt werden, dass dies eine Nebenwirkung des neuen URL-Security-Checks ist, allerdings können solche Äußerungen – wie oben erwähnt – auf Empfänger*innen verwirrend wirken. Alternativ könnte die Funktionalität des Clients angepasst werden, sodass eine entsprechende Security-Intervention nicht mehr angezeigt wird bzw. die Überprüfung der Signatur so angepasst wird, dass zunächst die Ziel-URL wieder in die E-Mail integriert und dann die Signatur überprüft wird. Weiterhin könnte der Security-Check des Mailservers auch so konfiguriert werden, dass die Ziel-URL bei digital signierten E-Mails nicht durch eine Weiterleitungs-URL ersetzt wird. Dies könnte bei Empfänger*innen aber erneut zu Verwirrung führen, weil diese teilweise E-Mails mit und teilweise ohne Weiterleitungs-URL vorfinden. Dies müsste daher in der Security-Awareness-Maßnahme im Rahmen der Einführung des neuen URL-Security-Checks erläutert werden.

Darüber hinaus ist das Einführen von Weiterleitungs-URLs durch den Security-Check am Mailserver nur möglich, wenn E-Mails nicht verschlüsselt sind. Sollen S/MIME oder PGP eingesetzt werden, um E-Mails zu verschlüsseln, dann könnte das Einsetzen von Weiterleitungs-URLs aufseiten der E-Mail-Clients[8] anstelle auf der des Mailservers Abhilfe schaffen. Dies hätte jedoch den Nachteil, dass nur besagter E-Mail-Client verwendet werden kann, der über die entsprechend notwendige Funktionalität verfügt. Diese Funktionalität müsste entweder durch den E-Mail-Client nativ unterstützt oder durch eine entsprechende Erweiterung, wie dies etwa bei Thunderbird und Outlook möglich ist,[9] ergänzt werden. Diese Möglichkeit besteht jedoch insbesondere für mobile E-Mail-Clients wie Apple-Mail oder Gmail für Android nicht, sodass der Nutzer mit solchen E-Mail-Clients nicht durch den URL-Security-Check geschützt, aber die menschliche Prüfung auch nicht behindert wäre.

2.4 URL-Security-Check erhält Passwort-Reset-Links

Bei vielen Online-Diensten ist es möglich Passwörter zurückzusetzen. Hierzu wird häufig eine E-Mail mit einem entsprechenden Passwort-Reset-Link verschickt. Klicken Empfänger*innen auf diesen Link so erhält der URL-Security-Check ebenfalls die URL um das Passwort zurückzusetzen. Die Betreiber*innen und/oder Entwickler*innen des URL-Security-Checks könnten sich so Zugriff auf die entsprechenden Konten verschaffen. Dies ist besonders kritisch, wenn die URL-Security-Check-Betreiber*innen/-Entwickler*innen andere Personen sind als die Mailserver-Betreiber*innen. Hier ist ein großes Maß an Vertrauen in die URL-Security-Check-Betreiber*innen/Entwickler*innen notwendig, dass diese lediglich die URL gegen Phishing-Datenbanken prüfen und nicht weiterverwenden bzw. speichern.

[8] E-Mail-Clients umfassen so wohl solche die über den Browser aufgerufen werden (Webmail), als auch lokale Programme.

[9] E-Mail-Client Thunderbird erweiterbar durch Add-ons. https://addons.thunderbird.net/de/thunderbird/ ; E-Mail Client Outlook erweiterbar durch Add-Ins. https://docs.microsoft.com/de-de/office/dev/add-ins/outlook/outlook-add-ins-overview .

2.5 Zwischenfazit

Aus technischer Sicht entstehen diverse Probleme beim Einsatz des herkömmlichen Mailserver-seitigen URL-Security-Checks. Zunächst sind die herkömmlichen Security-Awareness-Maßnahmen nicht mehr ausreichend und müssten, soweit dies möglich ist, angepasst werden. Zudem können Probleme bei der Weiterleitung und Beantwortung von E-Mails entstehen und der sinnvolle Einsatz von digital signierten E-Mails verhindert werden. Zuletzt ist die Ausführung des URL-Security-Checks aufseiten des Mailservers bei eingesetzter E-Mail-Verschlüsselung nicht möglich.

Um den eben genannten technischen Problemen entgegenzutreten und einen Mehrwert mittels URL-Security-Checks für die IT-Sicherheit bieten zu können, sollte der Einsatz wie folgt erfolgen: (1) die Ersetzung der URL sollte lokal im Client erfolgen, sodass es keine Probleme mit S/MIME bzw. PGP gibt und bei einer Weiterleitung bzw. Beantwortung der E-Mail nur noch die ursprüngliche Ziel-URL enthalten ist; (2) im E-Mail-Client sollten beide URLs angezeigt werden, wie es beispielsweise in Abbildung 2 gezeigt wird; (3) im Rahmen der Awareness-Maßnahmen sollte sichergestellt werden, dass die Empfänger*innen von E-Mails verstehen, dass sie nach wie vor angehalten sind, die Ziel-URL vor dem Klicken des Links zu prüfen.

3 Rechtliche Bewertung

Das Ersetzen einer URL hinter einem Link in einer E-Mail aus Gründen der IT-Sicherheit berührt sowohl den Schutz personenbezogener Daten als auch das Telekommunikationsgeheimnis. Durch die Nutzung könnte sogar gegen strafrechtliche Normen, insbesondere die §§ 206 und 303a StGB[10], verstoßen werden.[11] Die Frage der Zulässigkeit und die rechtlich zu betrachtenden Probleme richten sich wesentlich danach, in welchem Umfeld ein solches System eingesetzt wird. Hierbei ist insbesondere zwischen dem Einsatz für E-Mails im privaten Kontext und am Arbeitsplatz zu unterscheiden. Um eine inhaltliche Konzentration, auf die im Kontext von Weiterleitungs-URLs relevanten Probleme erreichen zu können, soll im Rahmen der rechtlichen Analyse von der Prämisse ausgegangen werden, dass die Funktionalität des eingesetzten URL-Security-Checks strikt auf das Erkennen von Links in E-Mails, die Umwandlung in Weiterleitungs-URLs sowie auf die Weiterleitung über den sicheren Link beschränkt ist. Von der Vornahme anderer Analysen soll somit ebensowenig ausgegangen werden, wie von einer missbräuchlichen Nutzung von Links zur Passworterneuerung. Für die Betrachtung soll zudem

[10] Danach ist die Verletzung des Fernmeldegeheimnisses unter den Voraussetzungen des § 206 StGB strafbar. Ebenso kann die Löschung, Unterdrückung, das Unbrauchbarmachen und Verändern von Daten gemäß § 303a StGB strafbar sein.

[11] Vgl. OLG Karlsruhe, Beschl. v. 10.01.2005, 1 Ws 152/04, Rn. 18 ff.; Heidrich/ Tschoepe, MMR 2004, 75, 76 ff.; 79 f.; Eckhardt in: Spindler/Schuster (Hrsg.), Recht der elektronischen Medien, 4. Aufl., 2019, §88 TKG, Rn. 67; Schröder in: Forgó/Helfrich/Schneider, Betrieblicher Datenschutz, 3. Aufl., 2019, Kap. 3, Rn. 39; Sassenberg/Mantz, BB 2013, 889, 892 f.

angenommen werden, dass die zuvor ausgesprochenen Empfehlungen zum optimalen Einsatz von URL-Security-Checks berücksichtigt werden, sodass dessen Durchführung das Potenzial für eine Verbesserung der IT-Sicherheit bietet. Problematisch ist dabei letztlich nur die Analyse und Veränderung der Links, da wir im Folgenden davon ausgehen, dass eine weitere Speicherung und Nutzung z.B. zur Auswertung von Angaben, wie wohin die ursprünglichen Links führten, nicht erfolgt. Vor diesem Hintergrund steht rechtlich somit nur in Frage, ob eine E-Mail von dem System auf das Vorhandensein von Links ausgewertet und solche durch Security Links ersetzt werden dürfen.

3.1 Einsatz im privaten Umfeld

Der Einsatz im privaten Kontext und die damit verbundene Beeinträchtigung des Telekommunikationsgeheimnisses und des Schutzes personenbezogener Daten kann durch die Einwilligung der*s E-Mail-Empfängers*in gerechtfertigt werden, die diese*r regelmäßig durch seine*ihre Zustimmung zu den Nutzungsbedingungen eines E-Mail-Providers oder -Clients erteilt. Erforderlich ist insoweit selbstverständlich, dass die Anforderungen an die wirksame Erteilung einer datenschutzrechtlichen Einwilligung gewahrt werden.[12] Neben den Rechten des*r Empfängers*in werden durch das Ersetzen eines Links jedoch auch die Rechte des*r Absenders*in, als anderem Teil der Kommunikation, betroffen. Dabei ist in Fällen, in denen eine Veröffentlichung von E-Mails ohne die Zustimmung des*r Absenders*in oder gar gegen seinen*ihren Willen erfolgt, anerkannt, dass sowohl eine Persönlichkeitsrechtsverletzung[13] als, bei ausreichender Schöpfungshöhe, auch eine Urheberrechtsverletzung[14] im Raum stehen kann. Hierfür wird im Wesentlichen auf die Interessenlagen in den anerkannten Fallkonstellationen der allgemeinen Zugänglichmachung von Briefen rekurriert.[15] Anders als in den gerichtlich entschiedenen Konstellationen findet im vorliegenden Fall jedoch keine Veröffentlichung der Inhalte und somit keine Zugänglichmachung zu einem unbestimmten Personenkreis statt. Dies lässt urheberrechtliche Bedenken ausscheiden. Zugleich wird das Interesse an der Vertraulichkeit des Wortes, aus dem sich die Verletzung des Persönlichkeitsrechts maßgeblich herleitet,[16] durch die Ersetzung der URL hinter dem Link, wenn überhaupt, nur unwesentlich beeinträchtigt. Der Ersetzung wird unabhängig davon, ob später überhaupt auf den Link geklickt werden wird, durchgeführt.

[12] Vgl. die Definition der Einwilligung in Art. 4 Nr. 11 DSGVO, wobei insbesondere die Freiwilligkeit und die Informiertheit der Einwilligung in der Praxis problematisch sein können (hierzu: Uecker, ZD 2019, 248; Klement in: Simitis/Hornung Spiecker gen. Döhmann, DSGVO, 2019, Art. 7 Rn. 48 f.; 73 ff.).

[13] LG Saarbrücken, Urt. v. 16.12.2011, 4 O 287/11; OLG Stuttgart, Urt. v. 10.11.2010, 4 U 96/10; LG Köln, Urt. v. 28.05.2008, 28 O 157/08; OLG Hamburg, Beschl. v. 04.02.2013; 7 W 5/13.

[14] Wissenschaftlicher Dienst des Bundestages, Die unbefugte Veröffentlichung privater Chat-Nachrichten Dritter, 2021, S. 12 f.

[15] LG Saarbrücken, Urt. v. 16.12.2011, 4 O 287/11; OLG Stuttgart, Urt. v. 10.11.2010, 4 U 96/10; LG Köln, Urt. v. 28.05.2008, 28 O 157/08.

[16] LG Saarbrücken, Urt. v. 16.12.2011, 4 O 287/11; OLG Hamburg, Beschl. v. 04.02.2013; 7 W 5/13.

Dies steigert die Tiefe des Eingriffs. Der tatsächliche Inhalt der Nachricht kann dabei andererseits jedoch ebenso wenig erfasst werden wie deren Absender*in. Selbst wenn eine Einwilligung des*r Absenders*in in eine automatische Ersetzung seiner*ihrer Links nicht erfolgt ist und auch nicht einfach angenommen werden kann, kann sie aufgrund einer vorzunehmenden Interessenabwägung erfolgen, wenn das Interesse des*r Empfängers*in, der seine*ihre Einwilligung erteilt hat, die Interessen des*r Absenders*in überwiegen.[17] Dies wird in der vorliegenden Konstellation regelmäßig der Fall sein. So findet das Ersetzen der URLs hinter Links automatisiert statt, wobei keine Analyse des eigentlichen Mailinhalts erfolgt. Die Vertraulichkeit der Kommunikation wird somit rein formal, aber nicht materiell berührt. Dem steht ein wichtiges Interesse des*r Empfängers*in gegenüber, der*die sich durch den Einsatz des Systems vor der Infiltration durch Schadsoftware schützen möchte. Trotz der unklaren Studienlage in Bezug auf die tatsächlichen Auswirkungen der Verwendung eines URL-Security-Checks (vgl. 1.2.1), kann dieser bei korrekter Anwendung (vgl. Empfehlungen in 1.2.5) einen echten Beitrag zur Verbesserung des IT-Sicherheit leisten. In weiterer Konsequenz können so sogar weitere IT-Systeme geschützt werden, indem keine Weiterverbreitung von Schadsoftware, z.B. über Adressbücher, erfolgt. Dem Interesse der IT-Sicherheit ist somit, nicht nur, aber insbesondere aus Empfänger*innensicht, einiger Stellenwert zuzuschreiben.

Datenschutzrechtlich kann ebenfalls auf die erteilte Einwilligung des*r Empfängers*in abgestellt werden. Der*die Absender*in einer E-Mail wird regelmäßig auch keine datenschutzrechtliche Einwilligung gegenüber dem Anbieter oder Client des*r E-Mail-Empfängers*in abgegeben haben. Vor dem Hintergrund der zu den Persönlichkeitsrechten des*r Absenders*in gemachten Ausführungen ergäbe sich eine datenschutzrechtliche Verarbeitungsgrundlage jedoch aus dem Überwiegen der berechtigten Interessen des*r Empfängers*in im Zusammenhang mit der IT-Sicherheit seiner*ihrer informationstechnischen Systeme gemäß Art. 6 Abs. 1 lit. f) DSGVO.

3.2 Einsatz am Arbeitsplatz

Es erscheint jedoch fraglich, ob der Einsatz eines entsprechenden Systems am Arbeitsplatz rechtlich zulässig ist, da zu seiner Funktion der Zugriff auf die E-Mail-Inhalte der Mitarbeitenden erforderlich ist.

Auch wenn dies zunehmend kritisiert wird,[18] besteht in Rechtsprechung und Literatur Einigkeit darüber, dass der*die Arbeitgeber*in den Inhalt einer E-Mail nur unter strengen Voraussetzungen einsehen darf. Dabei wird unterschieden, ob nur die dienstliche Nutzung von E-Mails am Arbeitsplatz erlaubt ist oder, ob der*die Arbeitnehmer*in das Mailsystem

[17] LG Saarbrücken, Urt. v. 16.12.2011, 4 O 287/11.

[18] Seifert in: Simitis/Hornung/Spiecker gen. Döhmann (Hrsg.), Datenschutzrecht, 2019, Art. 88 DSGVO, Rn. 150.

auch privat nutzen darf. [19] Bei der rein dienstlichen Nutzung darf der*die Arbeitgeber*in auf das Konto zugreifen, die Verbindungsdaten prüfen und sogar den Inhalt der dienstlichen E-Mails zur Kenntnis nehmen.[20] Ist dagegen auch die private Nutzung des Accounts erlaubt, gilt der*die Arbeitgeber*in als Anbieter*in von Telekommunikationsdiensten im Sinne des § 3 Nr. 1 TKG und ist gegenüber den Mitarbeiter*innen gemäß § 3 Abs. 1, 2 Nr. 2 TTDSG zur Wahrung des Fernmeldegeheimnisses verpflichtet.[21] In diesen Fällen ist es ihm*ihr verwehrt, auf den Mailverkehr seiner Mitarbeitenden zuzugreifen.[22] Dies gilt umso mehr als § 3 Abs. 3 TTDSG nunmehr festlegt, dass ein über das für den Schutz technischer Systeme erforderliche Maß hinausgehendes Kenntnisverschaffen von Inhalt und Umständen der Telekommunikation für Erbringer*innen von Telekommunikationsdienstleistungen verboten und die Verwendung von auf diesem Weg erlangten Informationen nur in engen Grenzen möglich ist. In der Literatur wird vorgeschlagen, die Mitarbeitenden aufzufordern, dienstliche und private E-Mails getrennt zu führen und zu speichern.[23] Allerdings ist eine Unterscheidung zwischen privater und geschäftlicher Korrespondenz ohne deren Mitwirkung nach wie vor nicht möglich.[24]

Der vorliegende Fall weist jedoch Parallelen zum Problem der Filterung von E-Mails durch Spam-Filter auf. In einem Kontext, in dem nur dienstliche E-Mails erlaubt sind, ist

[19] Seifert in: Simitis/Hornung/Spiecker gen. Döhmann (Hrsg.), Datenschutzrecht, 2019, Art. 88 DSGVO, Rn.148; Byers in: Weth/Herberger/Wächter/Sorge (Hrsg.), Daten- und Persönlichkeitsschutz im Arbeitsverhältnis, Kap. X., Rn. 21; Schröder in: Forgó/Helfrich/Schneider, Betrieblicher Datenschutz, 3. Aufl., 2019, Kap. 3, Rn. 39; Reichold, in Kiel/Lunk/Oetker, Münchener Handbuch zum Arbeitsrecht, Band I: Individualarbeitsrecht, 4. Aufl.,2018, §55, Rn. 35.

[20] Seifert in: Simitis/Hornung/Spiecker gen. Döhmann (Hrsg.), Datenschutzrecht, 2019, Art. 88 DSGVO, Rn. 149; Reichold, in Kiel/Lunk/Oetker, Münchener Handbuch zum Arbeitsrecht, Band I: Individualarbeitsrecht, 4. Aufl., 2018, §55, Rn. 35; Byers in: Weth/Herberger/Wächter/Sorge (Hrsg.), Daten- und Persönlichkeitsschutz im Arbeitsverhältnis, Kap. X., Rn. 22.

[21] Seifert in: Simitis/Hornung/Spiecker gen. Döhmann (Hrsg.), Datenschutzrecht, 2019, Art. 88 DSGVO, Rn. 150; Reichold, in Kiel/Lunk/Oetker, Münchener Handbuch zum Arbeitsrecht, Band I: Individualarbeitsrecht, 4. Aufl., 2018, §55, Rn. 35.

[22] Seifert in: Simitis/Hornung/Spiecker gen. Döhmann (Hrsg.), Datenschutzrecht, 2019, Art. 88 DSGVO, Rn. 150; Reichold, in Kiel/Lunk/Oetker, Münchener Handbuch zum Arbeitsrecht, Band I: Individualarbeitsrecht, 4. Aufl., 2018, §55, Rn. 35; Schröder in: Forgó/Helfrich/Schneider, Betrieblicher Datenschutz, 3. Aufl., 2019, Kap. 3, Rn. 39; Byers in: Weth/Herberger/Wächter/Sorge (Hrsg.), Daten- und Persönlichkeitsschutz im Arbeitsverhältnis, Kap. X., Rn. 21 f.

[23] Schröder in: Forgó/Helfrich/Schneider, Betrieblicher Datenschutz, 3. Aufl., 2019, Kap. 3, Rn. 42; Byers in: Weth/Herberger/Wächter/Sorge (Hrsg.), Daten- und Persönlichkeitsschutz im Arbeitsverhältnis, Kap. X., Rn. 22.

[24] Byers in: Weth/Herberger/Wächter/Sorge (Hrsg.), Daten- und Persönlichkeitsschutz im Arbeitsverhältnis, Kap. X., Rn. 22.

eine Filterung angesichts der oben beschriebenen Rechtslage unproblematisch.[25] Soweit aber auch privater Mailverkehr erlaubt ist, muss wiederum das Fernmeldegeheimnis beachtet werden.[26] Die sofortige Löschung jeder Spam-Mail wird daher als problematisch angesehen.[27] Lediglich die Löschung von E-Mails, die eine Gefahr für das IT-System des Unternehmens darstellen, weil sie Schadsoftware enthalten, ist aus Sicherheitsgründen zulässig.[28] In diesen Fällen sollten daher so genannte Quarantänelösungen vorgesehen werden, bei denen Spam-Mails in einen separaten Ordner verschoben werden und dort abrufbar sind, oder es besteht die Möglichkeit, den Betreff einer Nachricht mit dem Zusatz „Spam“ zu kennzeichnen.[29] Auf diese Weise werden zwar der Auffindeort der E-Mail oder aber ihr Betreff geändert, jedoch werden sie und ihr Inhalt weder unterdrückt, noch werden die in ihr enthaltenen Informationen insofern verändert oder verfälscht, als die ursprüngliche Nachricht nicht mehr transportiert würde, so dass auch der strafrechtliche Schutz des Fernmeldegeheimnisses nicht betroffen ist.[30]

Wendet man die betrachteten Grundsätze auf die vorliegende Fallgestaltung an, so kann man zu dem Schluss kommen, dass ihr Einsatz in Fällen, in denen nur eine dienstliche Nutzung am Arbeitsplatz zulässig ist, unproblematisch erscheint. Sind jedoch auch private E-Mails betroffen, muss der*die Arbeitgeber*in wiederum das Fernmeldegeheimnis wahren. Grundsätzlich dürfte er*sie daher nicht auf den Mailverkehr seiner*ihrer Angestellten zugreifen. Allerdings ist er*sie als Diensteanbieter*in im Sinne des § 3 Nr. 1 TKG auch nach § 165 Abs. 1 S. 1 TKG verpflichtet, angemessene technischen Vorkehrungen und sonstigen Maßnahmen zum Schutz des Fernmeldegeheimnisses und gegen Verletzungen des Schutzes personenbezogener Daten zu treffen.[31] Dies sieht auch § 3 Abs. 3 S. 1 TTDSG explizit als Ausnahme vom Fernmeldegeheimnis vor. Ein optimal

[25] Sassenberg/Mantz, Die (private) Emailnutzung im Unternehmen, BB 2013, 889, 892; Fuhlrott, in: Kramer, IT-Arbeitsrecht, 2. Aufl., 2019, Kap. B, Rn. 520 f.

[26] Sassenberg/Mantz, Die (private) Emailnutzung im Unternehmen, BB 2013, 889, 892; Fuhlrott, in: Kramer, IT-Arbeitsrecht, 2. Aufl., 2019, Kap. B, Rn. 520 f.

[27] Sassenberg/Mantz, Die (private) Emailnutzung im Unternehmen, BB 2013, 889, 892 f.; Eckhardt, in: Spindler/Schuster (Hrsg.), Recht der elektronischen Medien, 4. Aufl., 2019, §88 TKG, Rn. 62; Fuhlrott, in: Kramer, IT-Arbeitsrecht, 2. Aufl., 2019, Kap. B, Rn. 520; vgl. auch: OLG Karlsruhe, Beschl. v. 10.01.2005, 1 Ws 152/04, Rn. 21 ff.

[28] Sassenberg/Mantz, Die (private) Emailnutzung im Unternehmen, BB 2013, 889, 892; Sassenberg/Lammer, Zulässigkeit der Spam-Filterung im Unternehmen, DuD 2008, 461, 463; Heidrich/Schöpe, MMR 2004, 75, 78; Fuhlrott, in: Kramer, IT-Arbeitsrecht, 2. Aufl., 2019, Kap. B, Rn. 520 f.; OLG Karlsruhe, Beschl. v. 10.01.2005, 1 Ws 152/04, Rn. 25; VG Karlsruhe, Urt. v. 19.09.2007, 7 K 851/04, Rn. 23.

[29] Sassenberg/Mantz, Die (private) Emailnutzung im Unternehmen, BB 2013, 889, 892 f.

[30] Sassenberg/Mantz, Die (private) Emailnutzung im Unternehmen, BB 2013, 889, 892 f.; Heidrich/Tschoepe, Rechtsprobleme der E-Mail-Filterung, MMR 2004, 75, 78; OLG Karlsruhe, Beschl. v. 10.01.2005, 1 Ws 152/04, Rn. 22.

[31] Eckhardt, in: Geppert/Schütz (Hrsg.), BeckOK TKG, 4. Aufl., 2013, § 109, Rn. 15; Schommertz/Gerhardus, in: Scheuerle/Mayen, TKG, 3. Aufl., 2018, § 109, Rn. 3.

implementierter Security-Check hat das Potential vor schädlichen Links in E-Mails zu schützen, deren Folgen Datenverluste, die Infizierung ganzer IT-Systeme in Unternehmen und der Zugriff auf Informationen durch Unbefugte sind.

Problematisch erscheint im Rahmen dieser Schutzmaßnahme jedoch, dass der Inhalt einer E-Mail analysiert werden muss, um festzustellen, ob ein Link in der E-Mail enthalten ist. Es ist jedoch zu beachten, dass Spam-Filter auch den Inhalt (Body) der E-Mail heranziehen, um zu beurteilen, ob es sich um Spam handelt. Selbst dies erfolgt im vorliegenden Fall jedoch nicht, sodass weder Inhalt noch Absender*in analysiert, sondern nur Links identifiziert werden. Deren Inhalt, in Form der Weiterleitung zu einer bestimmten Zielseite, wird auch nicht so verändert, dass die durch sie transportierten Informationen verfälscht würden, da der ursprüngliche Link nicht gelöscht, sondern, wenn überhaupt nur durch ein Informationsfenster ergänzt wird. Damit bleibt auch der ursprüngliche Inhalt der E-Mail erhalten und wird nicht unterdrückt.

Ein weiterer kritischer Aspekt ist, dass nicht nur E-Mails mit schädlichen Links von der Analyse betroffen sind, sondern auch solche ohne Links oder mit Links, die nicht zu problematischen Seiten führen. Im Rahmen eines Spamfilters werden aber auch alle E-Mails, also auch die ungefährlichen, geprüft, um zwischen potentiell gefährlichen und nicht gefährlichen E-Mails zu unterscheiden. Wenn man vorher und ohne Prüfung einer E-Mail wüsste, ob sie potentiell gefährlich ist, bräuchte man keinen Filter. Eine Kategorisierung ist also nicht möglich, ohne jede E-Mail zu analysieren.

E-Mails enthalten in der Regel personenbezogene Daten, die bei der Auswertung auch vom Programm zumindest miterfasst werden. Es liegt daher auch eine Verarbeitung personenbezogener Daten im Sinne von Art. 2 Nr. 1, 2 DSGVO. Da das Programm am Arbeitsplatz eingesetzt werden soll, ist die Öffnungsklausel des Art. 88 Abs. 1 DSGVO auf den Beschäftigungskontext anwendbar. Vor diesem Hintergrund ist die maßgebliche Rechtsnorm daher § 26 BDSG. Solange keine Einwilligung des*r Arbeitnehmers*in vorliegt (vgl. § 26 Abs. 2 BDSG), muss die Verarbeitung der Daten daher erforderlich sein. Die Erforderlichkeit ist gegeben, wenn das Interesse an der Datenverarbeitung bei der Abwägung der widerstreitenden Interessen das Interesse an der Nichtverarbeitung der Daten überwiegt.[32] Im vorliegenden Zusammenhang steht also das Interesse des*r Arbeitsgebers*in an der Sicherheit und Integrität der informationstechnischen Systeme in seinem*ihrem Unternehmen dem Interesse der Arbeitnehmer*innen gegenüber, dass die sie betreffenden personenbezogen Daten nicht verarbeitet werden. Dies geschieht bei einem anwendungskonformen URL-Security-Check auch nicht, da sie weder als personenbezogene Daten erkannt, verstanden oder zugeordnet, noch gespeichert oder weiterverarbeitet werden. Vor diesem Hintergrund ist das Schutzbedürfnis des*r Arbeitnehmers*in deutlich geringer als das Interesse und die Pflicht des*r

[32] BT-Drs. 18/11325, S.97; Riesenhuber, in: Wolff/Brink (Hrsg.), BeckOK Dstenschutzrecht, 34. Ed. 01.11.2020, §26 BDSG, Rn. 113; Zöll, in: Taeger/Gabler (Hrsg.), DSGVO - BDSG, 3. Aufl., 2019, §26 BDSG, Rn. 23.

Arbeitgebers*in, die IT-Sicherheit seiner*ihrer Systeme mithilfe eines optimal implementierten URL-Security-Checks (siehe Empfehlungen in 1.2.5) zu gewährleisten.

Bei der Implementierung eines solchen Tools sollte der Betriebsrat beteiligt werden, da durch die Technik seine Rechte und Aufgaben berührt werden können, z.B. § 80 Abs. 1 Nr. 1, § 87 Abs. 1 Nr. 6 BetrVG. Da sich der Einsatz des Tools auf die Arbeitsabläufe am Arbeitsplatz der Arbeitnehmer*innen auswirkt, werden Informationspflichten des*r Arbeitgebers*in, § 81 Abs. 2 BetrVG,[33] sowie Vorschlagsrechte der Arbeitnehmer*innen, § 82 Abs. 1 BetrVG,[34] ausgelöst.

4 Fazit und Empfehlung

Die Nutzung von Weiterleitungs-URLs hat das Potenzial die IT-Sicherheit sowohl in Unternehmen als auch im privaten Bereich zu verbessern, sofern sie richtig umgesetzt und, im dienstlichen Kontext, in geeignete Security-Awareness-Maßnahmen eingebettet wird. Aus technischer Sicht sollte es den Empfänger*innen von E-Mails möglich sein, vor dem Klick auf den Link zu prüfen, ob die URL, die sie letztlich aufrufen wollen, plausibel und vertrauenswürdig ist. Hierzu bedarf es jedoch in den meisten Fällen einer entsprechenden Sensibilisierung durch Security-Awareness-Maßnahmen. Um die URL-Security-Checks durchführen zu können, ist es erforderlich eine Weiterleitungs-URL anstelle der Ziel-URL einzubinden. Dabei sollte weiterhin die Möglichkeit der Prüfung der Ziel-URL, sowie die vertrauenswürdige Verarbeitung anfallender Daten durch die URL-Weiterleitung gewährleistet werden. Auch ist für das Ersetzen der URL und die Weiterleitung die Verwendung eines internen Dienstes einem externen Dienstleister vorzuziehen. Wird ein externer Dienst verwendet, muss dieser insbesondere beim URL-Security-Check angesichts der potenziellen Einwirkungsmöglichkeiten besonders vertrauenswürdig und die Beschränkung der Funktionalität vertraglich geregelt sein.

Bei einer Weiterleitungs-URL kann zudem in geeigneter Weise die Ziel-URL im hinteren Teil lesbar enthalten sein. Dies ermöglicht in vielen Fällen weiterhin das Anwenden der Security-Awareness-Maßnahmen zur Prüfung der Ziel-URL, wobei Empfänger*innen den Aufbau der Weiterleitungs-URL für den verwendeten Dienst jedoch kennen und ihm vertrauen müssen. Zudem muss die URL vom E-Mail-Client in geeigneter Weise und ausreichender Länge angezeigt werden, sodass der Wer-Bereich der URL dargestellt wird. Dies ist jedoch nicht immer möglich. Daher ist für den URL-Security-Check eine Ersetzung seitens E-Mail-Client wie in Abschnitt 2.5 näher beschrieben zu empfehlen.

[33] Vgl. Kania, in: Müller-Glöge/Preis (Hrsg.), Erfurter Kommentar zum Arbeitsrecht, 21. Aufl, 2021, §81 BetrVG, Rn. 13; Werner, in: Rolfs/Giesen/Kreikebohm/Meßling/Udsching (Hrsg.), BeckOK Arbeitsrecht, 58. Ed., 01.12.2020, §81 BetrVG, Rn. 8.

[34] Vgl. Kania, in: Müller-Glöge/Preis (Hrsg.), Erfurter Kommentar zum Arbeitsrecht, 21. Aufl, 2021, §82 BetrVG, Rn. 3 f.; Werner, in: Rolfs/Giesen/Kreikebohm/Meßling/Udsching (Hrsg.), BeckOK Arbeitsrecht, 58. Ed., 01.12.2020, §81 BetrVG, Rn.1.

Eine andere Implementierung könnte zu einer Verringerung der Sicherheit führen, was eine neue rechtliche Bewertung erforderlich machen würde.

Rechtlich erscheint ein Einsatz von Weiterleitungs-URLs aus Gründen der IT-Sicherheit sowohl im privaten als auch im dienstlichen Umfeld möglich. Hierbei ist jedoch ebenso eine optimale Implementierung des Systems als auch seine Einbettung in geeignete Security-Awareness-Maßnahmen zu fordern. Zudem ist eine Beschränkung der Funktionalität des eingesetzten Tools auf die für Weiterleitungs-URLs erforderlichen Maßnahmen notwendig. Da in diesen Fällen von der potentiellen Verbesserung der IT-Sicherheit durch den Einsatz von Weiterleitungs-URLs auszugehen ist, überwiegt dann aber das Interesse am Einsatz der Technik die entgegenstehenden Rechte nicht einwilligender Betroffener.

Sicherheits- und Datenschutzanalyse der E-Mail-Server von Krankenhäusern und Kliniken

Eine praktische Analyse auf Datenschutzkonformität gemäß den Vorgaben der Datenschutzaufsichtsbehörden

Tim Wambach[1]

Abstract: Mit einer Orientierungshilfe formulierten 2021 die unabhängigen Datenschutzbehörden des Bundes und der Länder Mindestanforderungen für den technischen Schutz der Übermittlung personenbezogener Daten durch die E-Mail-Infrastruktur. Dabei wird, je nach Risiko, eine wirksame Verschlüsselung der Daten auf dem Übertragungsweg gefordert. Die hier vorgestellte Studie zeigt, dass nur 7 von 822 überprüften Mailserver die Anforderungen für normales Risiko der Orientierungshilfe erfüllen. Die Arbeit soll verdeutlichen, an welchen Stellen die derzeitige Praxis hinter den Erwartungen der Aufsichtsbehörden bleibt.

Keywords: Datenschutz, E-Mail, Sicherheit, TLS, Transportverschlüsselung, DSGVO

1 Einleitung

Im Mai 2021 hat die Konferenz der unabhängigen Datenschutzaufsichtsbehörden des Bundes und der Länder (DSK) eine Orientierungshilfe zum Schutz personenbezogener Daten bei der Übermittlung per E-Mail veröffentlicht[2]. Diese soll Schutzanforderungen konkretisieren, die bei Versand und Entgegennahme von E-Mail-Nachrichten durch Verantwortliche[3]zu berücksichtigen sind.

Grundsätzlich kann bei der Übertragung von E-Mails ein Schutz auf dem Transportweg durch Verschlüsselung erreicht werden. Dabei wird zwischen E-Mail-Client und E-Mail-Server des Senders, zwischen E-Mail-Server des Senders und E-Mail-Server Empfängers und E-Mail-Client des Empfängers und E-Mail-Server des Empfängers eine Transportverschlüsselung auf Basis von TLS eingesetzt. Die Orientierungshilfe der Aufsichtsbehörden fokussiert auf den Schutz zwischen den E-Mail-Servern des Senders

[1] Hochschule für Polizei und öffentliche Verwaltung Nordrhein-Westfalen, Abteilung Köln, Erna-Scheffler-Str. 4, 51103 Köln, tim.wambach@hspv.nrw.de

[2] „Maßnahmen zum Schutz personenbezogener Daten bei der Übermittlung per E-Mail“, abrufbar unter https://www.datenschutzkonferenz-online.de/media/oh/20210616_orientierungshilfe_e_mail_verschluesselung.pdf, zuletzt abgerufen am 06.03.2022.

[3] Entsprechend der Definition aus Art. 4 Nr. 7 DSGVO. Die Anforderung richten sich in gleicher Weise an Auftragsverarbeiter, die im Folgenden aus Gründen der Übersichtlichkeit nicht weiter genannt werden.

doi:10.18420/inf2022_50

und Empfängers (auch MTAs genannt). Sie spezifiziert Maßnahmen, die Verantwortliche auf Sender- und Empfängerseite berücksichtigen müssen, mit dem Ziel ein möglichst hohes Schutzniveau zu erreichen.

Es stellt sich die Frage, ob die Anforderungen der Aufsichtsbehörden, ein Jahr nach Veröffentlichung der Orientierungshilfe, in der Praxis eingehalten werden. Dies soll insbesondere bei einer Gruppe von Verantwortlichen überprüft werden, an die besonders hohe Erwartungen bezüglich des Schutzes personenbezogener Daten gestellt werden. So wurden 938 E-Mail-Server von Verantwortlichen aus dem Gesundheitswesen auf eine Umsetzung der Schutzmaßnahmen aus der Orientierungshilfe überprüft. Genauer handelt sich um E-Mail-Server von Krankenhäusern und Kliniken. Neben einer technischen Übersicht und einer Zusammenfassung der Forderungen der Aufsichtsbehörden werden die Ergebnisse dieser quantitativen Studie vorgestellt.

2 Technischer Ablauf beim E-Mail-Versand

Die Orientierungshilfe zielt auf die Schutzmaßnahmen auf dem Übertragungsweg zwischen dem E-Mail-Server des Senders und dem E-Mail-Server des Empfängers ab. Nicht berücksichtigt werden mögliche weitere Datenschutzanforderungen an Verantwortliche bzgl. Fragen der Aufbewahrung, Speicherbegrenzung oder Zweckbindung. In diesem Abschnitt sollen die technischen Hintergründe bei Versand und Empfang von E-Mails näher beschrieben werden.

Der Versand von E-Mail-Nachrichten erfolgt über SMTP[4]. Bei Versand einer E-Mail wird unter Verwendung einer E-Mail-Anwendung (Mail User Agent), wie beispielsweise Outlook, im ersten Schritt ein so genannter Mail Submission Agent (MSA) beauftragt die Nachricht zu übermitteln. Der MSA übergibt diese im zweiten Schritt an einen Mail Transmission Agent (MTA).

Im dritten Schritt wird der Nameserver der Empfängerseite bezüglich des zuständigen Mailservers befragt (MX-Eintrag) um im vierten Schritt über das SMTP-Protokoll die Nachricht an diesen zu übermitteln. Der MTA der Gegenseite übergibt die E-Mail dem Mail Delivery Agent (MDA) im fünften Schritt. Dieser bewahrt die E-Mail auf, bis der Empfangende im sechsten Schritt unter Verwendung der eigenen E-Mail-Anwendung (MUA) die E-Mail abruft.

Der Ablauf wird in Abb. 1 zusammengefasst, wobei die Pfeilrichtung die Richtung der Kontaktaufnahme zeigt. Es handelt sich um eine vereinfachte Darstellung – nicht berücksichtigt werden Schutzmaßnahmen zur SPAM-Prävention wie z. B. Blacklist-Anbieter oder der Einsatz von SPF (vgl. Abschnitt 3.4), für die weitere Abfragen und Kontaktaufnahmen erforderlich sind. Je nach Infrastruktur und eingesetzter Software

[4] Protokollbeschreibung des SMTP: Simple Mail Transfer Protocol abrufbar unter https://datatracker.ietf.org/doc/html/rfc5321 und zuletzt abgerufen am 06.03.2022.

werden die Rollen MSA/MTA bzw. MTA/MDA auch in einer Applikation zusammengefasst. Breite Anwendung finden hierbei die Serveranwendungen wie Postfix, Exim, qmail und Exchange.

Im Fokus der Orientierungshilfe der Aufsichtsbehörden steht eine Absicherung der Kommunikation zwischen den Verantwortlichen – also Schritt drei (DNS-Abfrage[5]) und vier (Übermittlung). Während die Schritte 1 und 2 sowie 5 und 6 üblicherweise in der eigenen Hand der jeweiligen Verantwortlichen oder dessen Auftragsverarbeiters liegen, stellt sich die Frage, wie die Übermittlung zwischen Sender und Empfänger gesichert werden kann. Das Ziel, in Bezug auf Vertraulichkeit, Integrität und Authentizität ein möglichst hohes Schutzniveau auf dem Transportweg zu erreichen, steht dabei jedoch in Konflikt mit dem Ziel, möglichst jede E-Mail-Nachricht entgegennehmen zu können (Verfügbarkeit).

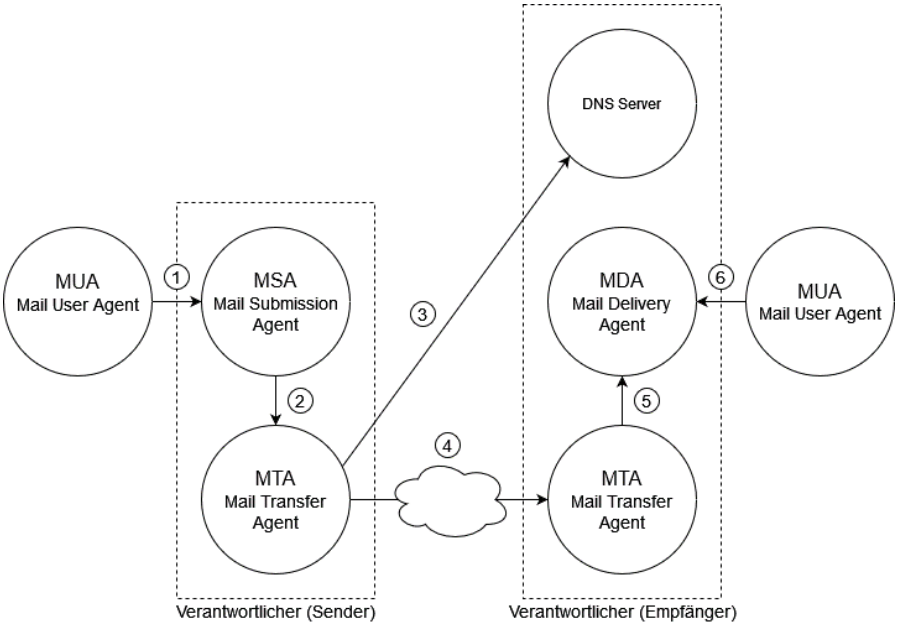


Abb. 1: Beteiligte Parteien bei Übertragung einer E-Mail

[5] DNS-Antwort kann möglicherweise auch aus einem Cache stammen. Dabei ist zu berücksichtigen, dass der Verantwortliche dies beeinflussen kann.

3 Bestehende Schutzmöglichkeiten

In diesem Abschnitt sollen die technischen Schutzmöglichkeiten kurz erläutert und bezüglich ihrer Schutzwirkung eingeordnet werden. Die Schutzmaßnahmen werden in Tab. 1 zusammengefasst.

Technologie	Schutzziel
(SSL/) TLS	Schutz der Vertraulichkeit und Authentizität des Zielservers bei Übermittlung.
MTA-STS[6]	Sicherstellung der Verschlüsselung, Schutz vor Downgrade-Attacken sowie begrenzter Schutz der Authentizität des Zielservers.
DANE[7]	Sicherstellung der Verschlüsselung und Authentizität des Zielservers, erfordert DNSSEC.
TLSA[8]	Bereitstellung von Zertifikatsinformationen durch einen DNS-Eintrag.
SPF[9]	Schutz vor gefälschten E-Mail-Absendern durch DNS-Einträge.
DKIM[10]	Schutz vor gefälschten E-Mail-Absendern durch digitale Signaturen des sendenden MTA.
DMARC[11]	Basiert auf SPF und DKIM und regelt den Umgang mit E-Mails bei Fehlern.
S/MIME, PGP, GPG	Ende-zu-Ende-Verschlüsselung sowie Authentizität des Absenders und Integrität der E-Mail-Nachricht durch Signaturen.

Tab. 1: Technologien und Schutzziele

[6] „SMTP MTA Strict Transport Security" (MTA-STS) vgl. RFC 8461.
[7] „SMTP Security via Opportunistic DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS)" vgl. RFC 7672.
[8] „The DNS-Based Authentication of Named Entities (DANE) Transport Layer Security (TLS) Protocol: TLSA" vgl. RFC 6698
[9] „Sender Policy Framework (SPF) for Authorizing Use of Domains in E-Mail, Version 1" vgl. RFC 4408.
[10] „DomainKeys Identified Mail (DKIM) Signatures" vgl. RFC 6376.
[11] „Domain-based Message Authentication, Reporting, and Conformance (DMARC)" vgl. RFC 7489.

3.1 Transportverschlüsselung durch TLS

Bei einer Transportverschlüsselung mittels TLS werden die Daten verschlüsselt, integritätsgeschützt und das Protokoll ermöglicht zusätzlich eine Authentifizierung und einen Schutz vor Replay-Angriffen.

Ursprünglich wurde SMTP als unverschlüsseltes Klartextprotokoll entwickelt. Sukzessive wurden Sicherheitsmaßnahmen hinzugefügt, um das Schutzniveau zu erhöhen und gleichzeitig abwärtskompatibel bleiben zu können. Sofern vom Server unterstützt, kann über den **STARTTLS** Befehl der E-Mail-Server angewiesen werden, eine SSL- bzw. TLS-geschützte Kommunikation auf Basis der bestehenden Verbindung zu ermöglichen. Auf diese Weise hatte der Verbindende die Wahl, auf welche Weise die Kommunikation fortgesetzt werden soll. Eine solche Wahlmöglichkeit führt zu einer **opportunistischen Transportverschlüsselung**: die Sicherheit orientiert sich an den Möglichkeiten des Senders und erlaubt zudem eine unverschlüsselte Übertragung, sofern STARTTLS nicht befohlen wird. Neben STARTTLS besteht auch die Möglichkeit direkt nach der Verbindungsaufnahme eine verschlüsselte Verbindung zu initiieren (Implicit TLS).

Der Sender kann auch den Einsatz einer Verschlüsselung erzwingen oder andernfalls die Kommunikation abbrechen und den Mailsender über die fehlgeschlagene Übermittlung informieren. Dieser Modus wird als **obligatorische Transportverschlüsselung** (auch enforce TLS oder mandatory TLS) bezeichnet[12]. Anbieter wie beispielsweise Posteo ermöglichen die Aktivierung von mandatory TLS und führen nur dann die Übertragung durch, wenn eine verschlüsselte Verbindung zustande gekommen ist.

Innerhalb einer TLS-gesicherten Verbindung soll der Einsatz von Zertifikaten vor Man-in-the-Middle Angriffen schützen. Ein solches Zertifikat enthält den Hostname des Servers und einen damit verbundenen öffentlichen Schlüssel. Der öffentliche Schlüssel wird im Zuge der Kommunikation verwendet und für die Entschlüsselung ist der Kenntnis des korrespondierenden privaten Schlüssels erforderlich. Auf diese Weise kann die Gegenseite prüfen, ob Kenntnis über den privaten Schlüssel besteht. Grundsätzlich können beide Kommunikationspartner solche Zertifikate zur gegenseitigen Authentisierung einsetzen, üblicherweise weist sich jedoch nur der Server (Empfänger) über ein Zertifikat aus. Der Client übermittelt die Daten erst dann, wenn an der Authentizität des Servers keinen Zweifel besteht.

Die Zertifikate werden üblicherweise von einer dritten Partei beglaubigt. Sofern eine solche Beglaubigung von einem Dritten fehlt, handelt es sich möglicherweise um ein **selbstsigniertes** (self signed) **Zertifikat**. Auch kann es vorkommen, dass das Zertifikat für eine andere Domain als die verwendete ausgestellt wurde. So wäre ein Zertifikat für www.example.com nicht gültig für mail.example.com. Zertifikate werden nur für einen Zeitraum erteilt und können deshalb abgelaufen oder zurückgerufen worden sein.

[12]Der Begriff „Mandatory TLS“ wird in RFC 7672 näher definiert, vgl. https://datatracker.ietf.org/doc/html/rfc7672 zuletzt abgerufen am 06.03.2022.

Je nach eingesetzter Version von SSL bzw. TLS besteht eine breite Auswahl an verschiedenen Kombinationen von Algorithmen, auch Cipher Suites genannt. Um einen möglichst hohen Schutz sicherzustellen, wird von der Nutzung alter SSL- bzw. TLS-Versionen abgeraten. Stand 2022 raten das BSI und die IANA zum ausschließlichen Einsatz von TLS in der Version 1.2 und 1.3 vgl. Tab. 2.

SSL 1.0	nicht final veröffentlicht
SSL2.0	nicht länger empfohlen seit 2011
SSL 3.0	nicht länger empfohlen seit 2015
SSL 3.1 / TLS 1.0	nicht länger empfohlen seit 2020, Empfehlung zu Deaktivierung wurde aufgrund der Covid19-Pandemie auf März 2021 verschoben
TLS 1.1	
TLS 1.2	derzeitiger Standard
TLS 1.3	

Tab. 2: SSL/TLS-Versionen und Empfehlungsstatus IANA

Die Version legt den grundsätzlichen Konfigurationsrahmen fest. Auch ist es möglich innerhalb der Versionen eine spezifische Auswahl der unterstützten Algorithmen zu treffen. Client und Server handeln auf Basis dieser Auswahl einen gemeinsam unterstützen Algorithmus aus. Schlägt dieser Aushandlungsprozess fehl, kann die Verbindung nicht aufgebaut werden.

Im Zuge des Aushandlungsprozesses könnte ein Angreifer versuchen die Aushandlung einer möglichst schwachen Verschlüsselung zu forcieren, um die Angriffsposition zu verbessern. Aus diesem Grund sollten ausschließlich starke Verschlüsselungen zum Einsatz kommen. In der technischen Richtlinie 02102-2[13] gibt das BSI eine Empfehlung zu den nutzbaren Cipher Suites ab.

[13] „Technische Richtlinie TR-02102-2 Kryptografische Verfahren: Empfehlungen und Schlüssellängen“ in der Version 2022-01, zuletzt abgerufen am 06.03.2022 unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR02102/BSI-TR-02102-2.pdf;jsessionid=663AE421E20CD6ADFE6BC5E124067D31.internet482?__blob=publicationFile&v=4

3.2 MTA-STS

Bei erstem Verbindungsaufbau ist unklar, ob vom Server eine verschlüsselte Verbindung unterstützt wird. MTA-STS (Mail Transfer Agent-Strict Transport Security), näher spezifiziert in RFC 8461[14], adressiert das Problem des Downgrade-Angriffs.

Dabei gibt sich ein Angreifer in beide Richtungen als die jeweilig andere Kommunikationspartei aus. Ein Angreifer, der in der Lage ist, die Kommunikation zwischen zwei Mailservern mitzuhören und zu modifizieren, kann das STARTTLS-Angebot vom empfangenden Mailserver entfernen. In diesem Fall würde der sendende E-Mail-Server nicht in den TLS-geschützten Modus wechseln.

Um dieses Problem zu beheben, wird im Domain Name System (DNS) neben dem MX-Eintrag zum E-Mail-Server auch ein Verweis auf eine Ressource eines Webserver platziert. Vom Webserver können weitere Daten (Policy) abgerufen werden, wie die Information, dass der E-Mail-Server über eine Verschlüsselung verfügt und angefragt werden kann. Somit kann der Client sich auf diesem Umweg versichern, ob das STARTTLS-Angebot vom Server erwartet werden kann. Ein Angreifer müsste in der Lage sein die HTTPS-geschützte Policy zu modifizieren.

MTA-STS adressiert allein das Problem des Downgrade-Angriffs, nicht die Manipulation durch Dritte auf dem Kommunikationsweg. Aus diesem Grund wird in der RFC 8461 explizit gefordert, dass das Zertifikat des Mailservers von einer vertrauenswürdigen Zertifizierungsstelle ausgestellt wurde und nicht abgelaufen ist.

3.3 DANE

Um ein von der PKI losgelöstes Konzept zur zuverlässigen Verteilung von Zertifikaten zu erreichen kann ein DNS-basiertes Verteilungssystem für Zertifikate genutzt werden: DNS-based Authentication of Named Entities (kurz: DANE). Es ermöglicht die Überprüfung des Zertifikats des Zielservers ohne gemeinsames Wurzelzertifikat. Das dafür notwendige Schlüsselmaterial wird hierfür im Domain Name System (DNS) platziert: diese Einträge werden TLSA-Einträge genannt und ermöglichen die Zertifikatsüberprüfung. Um die Vertrauenswürdigkeit der Antworten sicherzustellen, müssen Antworten über DNSSEC (Domain Name System Security Extensions) gesichert sein. Wurde DNSSEC etabliert, ist ein Domaininhaber in der Lage Zertifikate für die eigene Domäne selbst auszustellen.

Die Nutzung von DANE setzt somit eine Unterstützung von DNSSEC voraus, um vor einer Manipulation der Zertifikatsinformationen durch Dritte zu schützen. Dies sowohl

[14] „SMTP MTA Strict Transport Security (MTA-STS)“, abrufbar unter https://datatracker.ietf.org/doc/html/rfc8461 und zuletzt abgerufen am 06.03.2022.

beim Empfänger, der Schlüsselmaterial im DNS ablegt, als auch beim Sender, der diese Informationen beziehen und zu validieren in der Lage sein muss.

3.4 DKIM, SPF und DMARC

Auch ohne Sicherung durch DNSSEC kann ein Eintrag im DNS die Sicherheit erhöhen. Ein **SPF**-Eintrag im DNS (vom Typ TXT seit 2014) soll vor gefälschten E-Mail-Absendern schützen. Über diesen Eintrag kann vom empfangenden Server geprüft werden, ob die IP-Adresse oder der Hostname des Servers, der eine E-Mail übertragen möchte, auch zur Übermittlung von E-Mails berechtigt ist. Auf diese Weise findet eine eingeschränkte Authentifizierung des sendenden E-Mail-Servers statt.

Auch ist es möglich, Schlüsselmaterial im DNS zu platzieren, um die Authentizität einer eingehenden E-Mail-Nachricht zu überprüfen. Dabei unterschreibt der sendende MTA die E-Mail mit einer digitalen Signatur und übermittelt diese als **DKIM**-Signatur in den Header-Daten der E-Mail. Der Zielserver kann auf Basis des im DNS abgelegten Schlüsselmaterials den Ursprung (im Sinne des sendenden Servers) gegenprüfen. DKIM soll somit das Fälschen von E-Mail-Absendern erschweren, da nur der sendende Mailserver zur Ausstellung einer gültigen Signatur in der Lage ist.

DMARC (Domain-based Message authentication, Reporting and Conformance) baut auf den Techniken SPF und DKIM auf und regelt darüber hinaus, wie der empfangende Mailserver im Falle eines Fehlers reagieren soll. Gemeinsam sollen diese Technologien die Authentizität des Absenders sicherstellen und indirekt vor SPAM-E-Mails schützen.

3.5 Ende-zu-Ende-Verschlüsselung

Eine Verschlüsselung, die beim Sender startet und bis zum Empfänger ununterbrochen bestehen bleibt, wird als Ende-zu-Ende-Verschlüsselung bezeichnet. Auf diese Weise soll insbesondere die Vertraulichkeit des Inhalts sichergestellt sein. Sofern Schlüsselmaterial vorhanden ist, kann über digitale Signaturen auch die Integrität und Authentizität gesichert werden. S/MIME und PGP bzw. GPG ermöglichen einen solchen Schutz.

4 Orientierungshilfe der DSK

Mit der Orientierungshilfe „Maßnahmen zum Schutz personenbezogener Daten bei der Übermittlung per E-Mail“ mit Stand vom 16.06.2021 haben die Aufsichtsbehörden ihre Anforderungen an E-Mail-Systeme gestellt, die personenbezogene Daten übermitteln. Je nach Risiko werden technische Anforderungen an beide Seiten, den Sender und den Empfänger, gestellt. Auch wenn die Verantwortung für die Übertragung beim Sender liegt, werden Voraussetzungen für den sicheren Empfang gefordert. An dieser Stelle sollen nur die für diese Analyse wesentlichen Informationen zusammengefasst werden.

4.1 Obligatorische- und qualifizierte Transportverschlüsselung

Die Orientierungshilfe unterscheidet zwischen der obligatorischen und der qualifizierten Transportverschlüsselung. Bei einer obligatorischen Transportverschlüsselung soll die unverschlüsselte Übertragung ausgeschlossen werden. Sofern der Sender oder Empfänger keine Verschlüsselung unterstützen, muss die Verbindung abgebrochen werden.

Bei einer qualifizierten Transportverschlüsselung wird auf die Anforderungen der Technischen Richtlinie BSI TR-02102-2 verwiesen. Dabei soll die IP-Adresse des Zielservers über DNSSEC validiert werden. Sofern die Gegenseite zertifikatsbasiert authentisiert wird, wird die Vertrauenswürdigkeit über ein gemeinsames Wurzelzertifikat oder via DANE publizierten Vertrauensanker überprüft.

4.2 Risiko

Die Orientierungshilfe unterscheidet zwischen zwei Arten von Risiken: normales und hohes Risiko für die Rechte und Freiheiten natürlicher Personen. Zur Einstufung von Risiken wird auf das Kurzpapier Nr. 18 der unabhängigen Datenschutzbehörden des Bundes und der Länder „Risiko für die Rechte und Freiheiten natürlicher Personen“ verwiesen[15].

Bei normalem Risiko wird gefordert[16], dass der empfangende Server mindestens

- den Aufbau von TLS-Verbindungen ermöglichen muss und
- hierbei ausschließlich die in der BSI TR 02102-2 aufgeführten Algorithmen verwenden darf.

Innerhalb dieses Rahmens sollte der Verantwortliche für Verschlüsselung und Authentifizierung ein möglichst breites Spektrum an Algorithmen anbieten. Bei normalem Risiko wird für den Sender eine obligatorische Transportverschlüsselung gefordert und empfohlen, sich an der TR 03108-1[17] zu orientieren. Um Authentizität und Integrität zu schützen, sollten DKIM-Signaturen zum Einsatz kommen und Fehler entsprechend den Festlegungen des Absenders (DMARC) behandeln.

---

[15] Kurzpapier zur Risikoeinstufung, abrufbar unter https://www.datenschutzkonferenz-online.de/media/kp/dsk_kpnr_18.pdf zuletzt abgerufen am 06.03.2022.

[16] Abschnitt „4.1.1 Verpflichtungen bei normalen Risiken“ https://www.datenschutzkonferenz-online.de/media/oh/20210616_orientierungshilfe_e_mail_verschluesselung.pdf zuletzt abgerufen am 06.03.2022.

[17] Secure E-Mail-Transport, verfügbar unter https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR03108/TR03108-1.pdf;jsessionid=A5F52626B2BD0E2B8207AC38E7F5BDE1.internet482?__blob=publicationFile&v=2 in der Version vom 17.02.2022.

Bei einem gezielten Empfang von E-Mails, deren Verlust der Vertraulichkeit ein hohes Risiko bedeuten, muss

- sowohl eine qualifizierte Transportverschlüsselung,
- als auch der Empfang von Ende-zu-Ende-verschlüsselten Nachrichten ermöglicht werden.

Sofern bei Verlust der Integrität ein hohes Risiko zu erwarten ist, muss ferner auch der Einsatz von digitalen Signaturen ermöglicht werden. Beim Versand wird offengelassen, wie qualifizierte Transportverschlüsselung und Ende-zu-Ende-Verschlüsselung zu kombinieren sind, um ein geeignetes Schutzniveau sicherzustellen.

Verantwortliche, die gemäß § 203 StGB zur Geheimhaltung von Kommunikationsinhalten verpflichtet sind, müssen die Anforderungen von normalem und hohem Risiko erfüllen und können darüber hinaus noch weitere Sicherheitsmaßnahmen ergreifen.

5 Analyse des aktuellen Stands

5.1 Design der Analyse

Ziel der quantitativen Analyse ist ein Vergleich der Vorgabe der Aufsichtsbehörden mit den derzeitigen Konfigurationen in der Serverlandschaft. Derartige Analysen von Mail-Servern im Internet wurden in der Vergangenheit schon häufiger durchgeführt[18].

Grundsätzlich lässt sich von außen nicht abschätzen, welche Informationen über einen E-Mail-Server ausgetauscht werden. Anhand der Zielgruppe der Analyse, Mailserver von Krankenhäusern und Kliniken, ist zu erwarten, dass personenbezogene Daten ein relevanter Teil davon sind. Auch wenn zu vermuten ist, dass eine umfassende Risikoabschätzung bzgl. der möglichen Gefahren beim E-Mail-Versand nicht immer durchgeführt wird, kann und soll nicht pauschal von einem hohen Risiko ausgegangen werden. Aus diesem Grund soll sich die Analyse nur an den Anforderungen der Aufsichtsbehörden für **normalem Risiko** orientieren, welches jedoch das Mindestmaß darstellt.

Ferner fokussiert sich die Analyse an den technischen Möglichkeiten, die von außen überprüfbar sind. Dies umfasst die Unterstützung der Transportverschlüsselung durch TLS (STARTTLS sowie Implicit TLS), die Bereitstellung eines gültigen Zertifikates, die

[18] Wilfried Mayer, Aaron Zauner, Martin Schmiedecker, Markus Huber: No Need for Black Chambers: Testing TLS in the E-mail Ecosystem at Large. ARES 2016: 10-20, abrufbar unter https://publications.sba-research.org/publications/scanTLS.pdf

Unterstützung von MTA-STS und DANE sowie die unterstützen Verschlüsselungsalgorithmen (Cipher Suites).

5.2 Werkzeuge

Die Analyse setzt darauf, möglichst erprobte Werkzeuge einzusetzen. Diese mussten in Teilen jedoch angepasst und erweitert werden, um eine quantitative Analyse zu ermöglichen. Zum Einsatz gekommen sind:

- die Prüfwebseite https://www.checktls.com/ ermöglicht eine Überprüfung der Mailserver in Bezug auf Verschlüsselung, Zertifikate sowie (sofern zusätzlich aktiviert) MTA-STS und DANE.
- Über das Python Modul sslyze wurden die Mailserver bzgl. der verfügbaren Protokolle und unterstützten Verschlüsselungen analysiert.
- Das zu OpenSSL zugehörige s_client-Werkzeug wurde genutzt, um zu prüfen, welche TLS-Protokolle verfügbar sind.
- Über das Python Modul dnspython wurden DNS-Einträge und die Unterstützung von DNSSEC ermittelt.

5.3 Ethische Aspekte und Einschränkungen

An dieser Stelle sollen auch die ethischen Aspekte einer solchen Analyse berücksichtigt werden. Protokollbedingt erfordert die Ermittlung der unterstützen Algorithmen eine Vielzahl an Verbindungsversuchen. Dabei bietet der Client bei jedem Verbindungsaufbau nur einen Algorithmus an und prüft, ob der Server diesen akzeptiert. Dabei fallen ca. 0,5 Megabyte Datenvolumen pro Server an. Aus diesem Grund wurde das Modul sslyze so modifiziert, die punktuelle Last möglichst gering zu halten, indem die Analyse auf eine längere Zeitspanne ausgedehnt wird. Trotz dieser Maßnahmen wurde beobachtet, dass einige Mailserver nach mehreren fehlgeschlagenen Verbindungsversuchen weitere Tests unterbinden. Aus diesem Grund muss davon ausgegangen werden, dass die Liste der unterstützen Cipher Suites nicht vollständig ist und ist als Untergrenze anzusehen.

Um weitgehende Vollständigkeit bzgl. der unterstützen Protokolle sicherzustellen, wurde die Ergebnisse mit openssl s_client erneut geprüft und ergänzt – dabei ist jeweils nur ein Verbindungsaufbau pro Protokoll erforderlich.

Sofern nicht explizit anders erwähnt, wurde nicht geprüft, ob die E-Mail-Adresse tatsächlich verwendet bzw. auf dem Webauftritt publiziert wurden. Auf ein solches Crawling von E-Mail-Adressen vom Internetauftritt wurde aufgrund der damit verbundenen Serverbelastung verzichtet.

5.4 Testmenge und -umfang

Ausganspunkt der Analyse war eine Testmenge von 3415 Adressen von Krankenhäusern und Kliniken aus verschiedenen öffentlich zugänglichen Quellen (kliniken.de, weisse-liste.de, Verzeichnisse von Krankenkassen).

Diese wurde dahingehend geprüft, ob die Webseiten noch erreichbar sind sowie Mehrfacheinträge entfernt. Für die verbleibenden 1810 Webauftritte wurde über das Domain Name System (DNS) der zugehörige Mailserver abgefragt – sofern ein Eintrag verfügbar war. Anschließend wurden Einträge zusammengefasst, die auf den gleichen Mailserver verweisen. Dies ist insbesondere bei größeren Klinikgruppen der Fall. Abschließend wurde getestet, ob ein Verbindungsaufbau zum Mailserver möglich ist, um veraltete Einträge auszuschließen. Entfernt wurden ebenfalls jene Kliniken, die Microsoft als Cloud-Dienstleister nutzen (121 Einrichtungen) – dies wird in Abschnitt „Microsoft als Dienstleister“ thematisiert.

Somit verblieben 936 E-Mail-Server, die von 1687 Einrichtungen genutzt werden bzw. deren Webauftritt über 1687 verschiedenen URLs erreichbar sind. Die Analysen fanden im Zeitraum vom 20.02.2022 bis 06.03.2022 statt.

6 Ergebnisse

Die Analyse ergab, dass 33 der 936 E-Mail-Server keine verschlüsselte Kommunikation ermöglichten. Der STARTTLS-Befehl wurde nicht angeboten und ein direkter verschlüsselter Verbindungsaufbau ist fehlgeschlagen. Eine daran anschließende stichprobenartige Betrachtung der Internetauftritte der Verantwortlichen zeigte jedoch, dass die E-Mail-Server in Verwendung zu sein scheinen.

In den verbleibenden 903 Mailservern, die eine Verschlüsselung unterstützen, wurden

- die Zertifikate,
- unterstützten SSL/TLS-Protokolle,
- unterstützten Verschlüsselungsalgorithmen,
- MTA-STS und
- DANE-Unterstützung

überprüft. Im Zuge der tiefergehenden Analysen ergaben sich in 81 der 936 Fälle Mailserver, die auf ein oder mehrere Analysewerkzeuge nicht reagiert haben. Dies kann durch vorgeschaltete Sicherheitsmaßnahmen (Firewalls, Gateways) begründet sein.

6.1 Überprüfung der Zertifikate

Neben den 33 E-Mail-Servern, die keine Verschlüsselung unterstützen, ergab sich, dass in 378 Fällen (40 %) die verwendeten Zertifikate nicht zur Authentifizierung genutzt werden können. Die Gründe hierfür sind, dass die Zertifikate selbstsigniert oder nicht für den Hostnamen des Mailservers ausgestellt waren.

Somit wird in diesen Fällen zwar eine Verschlüsselung ermöglicht, jedoch kann der Sender die Gegenseite vor der Übermittlung nicht authentifizieren. In den verbleibenden 525 Fällen (56 %) war mindestens von einem Mailserver ein gültiges Zertifikat verfügbar.

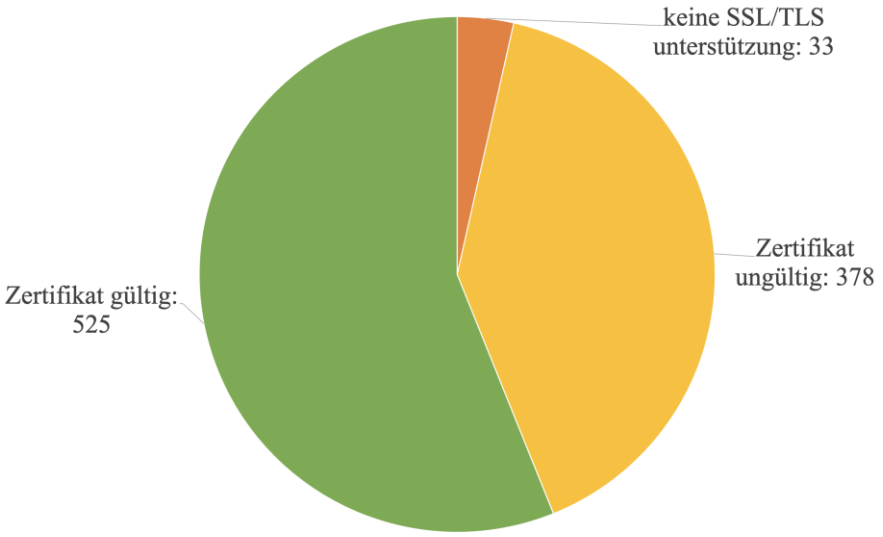


Abb. 2: Beteiligte Parteien bei Übertragung einer E-Mail

6.2 Unterstützte SSL/TLS-Protokolle

Wie bereits erwähnt waren hier 822 Testergebnisse aussagekräftig. In einem Fall wurde noch SSL in der Version 2 unterstützt. Bei 21 Fällen war noch eine Unterstützung von SSL in der Version 3.0 vorhanden. TLS in der Version 1.0 und 1.1 waren, entgegen der Empfehlung vom BSI, noch in 71 % bzw. 83 % unterstützt. Nahezu alle Mailserver haben eine Unterstützung von TLS in der Version 1.2 angeboten. Nur 27 % der Mailserver haben bereits eine Unterstützung von TLS in der Version 1.3.

In zwei Fällen war keine Unterstützung von TLS v1.2 und TLS v1.3 verfügbar – diese Ergebnisse wurden nochmal manuell validiert bestätigt – TLS v1.2 wurde im Jahr 2008 veröffentlicht. Somit konnten diese Server nur mittels TLS v1.1 und darunter erreicht werden.

Unterstützte Protokolle	Anzahl	Anteil	Alleinige Unterstützung
SSL v2.0	1	0,12 %	2: 0,24 %
SSL v3.0	24	2,92 %	
TLS v1.0	590	71,78 %	
TLS v1.1	689	83,82 %	
TLS v1.2	820	99,76 %	126: 15,33 %
TLS v1.3	227	27,62 %	
Gesamt	822		

Tab. 3: Liste der unterstützten Protokolle

In Tab. 1 werden die Ergebnisse zusammengefasst. Zu sehen sind die Anzahl der Mailserver, welche die Protokolle unterstützt haben und deren Anteil an der Gesamtmenge. In der letzten Spalte werden die Ergebnisse dahingehend zusammengefasst, wie viele Mailserver ausschließlich nicht-empfohlene (2) oder empfohlene (126) Versionen akzeptieren. Aus der Differenz zu 822 ergibt sich, dass 696 Server (ca. 84 %) nicht-empfohlene Versionen akzeptieren.

6.3 Unterstützte Algorithmen

Die Analyse der unterstützten Algorithmen bei den 822 Mailservern ergab, dass nicht empfohlene Algorithmen noch eine breite Anwendung finden. Die zwei am häufigsten eingesetzten Cipher Suiten TLS_RSA_WITH_AES_256_CBC SHA bzw. TLS_RSA_WITH_AES_128_CBC_SHA, die von 93 % der Sever unterstützt werden, gelten als schwache Verfahren. Dies ist u.a. mit einer seit 2013 bekannten Schwachstelle im CBC-Modus begründet[19].

Ein Vergleich der unterstützten Algorithmen mit den Empfehlungen des BSI[20] zeigte, dass in 7 der 822 Fällen die Mailserver ausschließlich vom BSI empfohlene Algorithmen

[19] vgl. Rizzo & Duong in Practical Padding Oracle Attacks. WOOT 2010.
[20] Technische Richtlinie TR-02102-2 Kryptografische Verfahren: Empfehlungen und Schlüssellängen in der Version 2022-01, zuletzt abgerufen am 06.03.2022 unter

unterstützt werden. In den verbleibenden 815 Fällen wurden Algorithmen unterstützt, die nicht empfohlen sind.

Bei einem Vergleich mit den Empfehlungen der IANA zeigten sich nur in 10 der 822 Fälle Mailserver, die ausschließlich von der IANA empfohlene Algorithmen einsetzen. In den verbliebenen 812 Fällen wird mindestens ein Algorithmus verwendet, der von der IANA nicht empfohlen wird.

Werden die Empfehlungen von IANA und BSI zusammengefasst entsprechen 14 der 822 E-Mail-Server den Vorgaben – die verbleibenden 808 Server verwenden mindestens einen nicht empfohlenen Algorithmus.

Besonders bemerkenswert ist eine Unterstützung von EXPORT Cipher Suiten in zwei Fällen. Dabei handelt es für den Zweck des Exports absichtlich geschwächte Verschlüsselungen und gelten als Relikt bzw. sind nicht länger Bestandteil von TLS seit Version 1.2. Im Jahr 2015 wurden Schwachstellen aufgrund dieser Algorithmen unter dem Akronym FREAK bekannt[21].

Kritisch ist ebenso, dass 9 % der Server die Stromchiffre RC4 unterstützen, die seit 2015 nicht länger eingesetzt werden sollte[22]. Auch die Unterstützung von 3DES bei 22 % der Server muss als Risiko betrachtet werden. Eine vollständige Übersicht der unterstützen Algorithmen ist in Tab. 4im Anhang zu finden.

6.4 MTA-STS

Die Unterstützung von MTA-STS wurde über die Überprüfung eines entsprechenden DNS-Eintrags geprüft. Unabhängig von dem Ergebnis wurde ebenso versucht die Policy über HTTPS abzurufen. Es zeigte sich, dass nur vier der 822 Mailserver eine entsprechende Policy anbieten, wobei einer davon einen fehlerhaften Inhalt aufwies.

6.5 DANE

Insgesamt 24 der 822 Mailserver wiesen eine Unterstützung von DANE auf (2 %). Dabei konnte der DNS-Eintrag über DNSSEC validiert und der öffentliche Schlüssel zur Überprüfung des Zertifikates abgerufen werden. Inwieweit DANE tatsächlich von den E-Mail-Servern beim Versand genutzt bzw. validiert wird, kann nicht ermittelt werden.

---

https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/TechnischeRichtlinien/TR02102/BSI-TR-02102-2.pdf

[21] Der FREAK-Angriff, abrufbar unter https://blog.cryptographyengineering.com/2015/03/03/attack-of-week-freak-or-factoring-nsa/ und zuletzt abgerufen am 06.03.2022.

[22] „Prohibiting RC4 Cipher Suites", RFC 7465, abrufbar unter https://datatracker.ietf.org/doc/html/rfc7465.

6.6 Microsoft als Dienstleister

Bei 121 Einrichtungen wurden die Ergebnisse vor Start der Analyse zusammengefasst und sind nicht Teil der 938 zuvor analysierten Mailserver. In diesen 121 Fällen wird als Dienstleister auf einen Service von Microsoft zurückgegriffen bzw. auf eine Microsoft-Domain verwiesen. Die Hostnamen der Mailserver entsprechen alle dem Aufbau „<domain>.mail.protection.outlook.com“. Da es sich um eine höhere Quantität handelt wurden diese aus der Gesamtbetrachtung herausgelöst.

Die Liste der gesichteten Cipher Suiten kann in Tab. 5 im Anhang eingesehen werden. Der Mailserver war ausschließlich über TLS in der Version 1.2 ansprechbar – eine Unterstützung von TLS 1.3 oder älteren Protokollen wurde nicht festgestellt.

6.7 Zusammenfassung der Ergebnisse

Entsprechend der Orientierungshilfe der Aufsichtsbehörden muss für normales Risiko eine TLS-geschützte Verbindung genutzt werden, wobei ausschließlich die vom BSI empfohlenen Algorithmen zum Einsatz kommen sollen. Die Analyse zeigt, dass dies nur in 7 der 822 Mailserver tatsächlich vorliegt. Ferner muss die Frage diskutiert werden, inwieweit von einem Schutz durch TLS ausgegangen werden kann, wenn das Zertifikat des E-Mail-Servers bei 40 % nicht geprüft werden kann.

In der technischen Richtlinie des BSI wird darauf hingewiesen, dass eine fehlende Empfehlung nicht grundsätzlich Unsicherheit bedeutet. Die Unterstützung von veralteten TLS-Protokollen bei 84 % der analysierten Mailserver ist allerdings als kritisch zu werten, da sich hier auch langfristig die Frage nach technischer Unterstützung und Sicherheitsupdates stellen würde.

Durch einen Vergleich mit den Empfehlungen der IANA zeigt sich, dass auch hier nur wenige Server (10) die Empfehlungen erfüllen. Sofern die Empfehlungen von BSI und IANA zusammengefasst werden, würden auch nur 14 Server diese einhalten.

Die geringe Abdeckung von DANE (2 %) und die Tatsache, dass in 40 % kein gültiges Zertifikat vom Mailserver ausgewiesen wird, deutet darauf hin, dass eine Umsetzung einer qualifizierten Transportverschlüsselung nur von einem kleinen Teil der Verantwortlichen angestrebt wird.

7 Diskussion

Die Analysen zeigen, dass die Forderungen der DSK in Bezug auf die eingesetzten Algorithmen bei den betrachteten E-Mail-Servern bisher nur wenig Berücksichtigung finden. Selbst für ein normales Risiko wird das angestrebte Schutzniveau nicht erreicht.

Die Unterstützung von nicht länger empfohlenen bzw. unterstützen Protokollen (SSL 2.0 bis TLS 1.1) bei mehr als 4 von 5 Servern ist als kritisch zu werten. Zwar kann der Wunsch nach einer möglichst breiten Kompatibilität mit älteren E-Mail-Servern aus Sicht der Betreiber nachvollzogen werden, allerdings ist TLS 1.2 bereits seit 14 Jahren veröffentlicht und es zeigten sich nur zwei Server, die dieses Protokoll nicht unterstützen. Als Betreiber wäre es in diesem Fall unkritisch die Unterstützung älterer Protokolle zu deaktivieren, was aus Sicherheitsgründen unmittelbar geschehen sollte.

Da 40 % der betrachteten Anbieter kein gültiges Zertifikat vorweisen, muss vermutet werden, dass dieses beim Versand auch nicht berücksichtigt wird. Es stellt sich jedoch die Frage, weshalb die E-Mail-Sicherheit hier so deutlich unterhalb der Möglichkeiten bleibt.

Die Ergebnisse vermitteln den Eindruck, dass im Zielkonflikt zwischen Vertraulichkeit und Authentizität auf der einen Seite und Verfügbarkeit auf der anderen Seite, die Verfügbarkeit derzeit einen höheren Stellenwert einnimmt. So ist zu vermuten, dass aus Angst vor Kompatibilitätsverlust mit älteren E-Mail-Servern eine möglichst breite Auswahl an Protokollen und Algorithmen angeboten wird.

Ähnliche Probleme konnten noch vor wenigen Jahren beim Wechsel von HTTP nach HTTPS beobachtet werden. Die Ankündigung einiger Browserhersteller, dass Webseiten ohne Verschlüsselung nur noch eingeschränkt verfügbar sein werden (Warnhinweise, insbesondere bei Passworteingabe), führte der äußeren Beobachtung nach zu einem schnellen Umdenken. Beim E-Mail-Verkehr könnten solche Änderungen von den führenden MTA-Softwareanbietern stärker forciert werden.

8 Anhang

Beschreibung	Sichtungen	Anteil	BSI	IANA
TLS_RSA_WITH_AES_256_CBC_SHA	709	93,78%	nein	nein
TLS_RSA_WITH_AES_128_CBC_SHA	707	93,52%	nein	nein
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	698	92,33%	nein	nein
TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384	697	92,20%	ja	nein
TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384	696	92,06%	ja	ja
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256	696	92,06%	ja	nein
TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	693	91,67%	ja	ja
TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA	691	91,40%	nein	nein
TLS_RSA_WITH_AES_256_CBC_SHA256	678	89,68%	nein	nein
TLS_RSA_WITH_AES_128_CBC_SHA256	670	88,62%	nein	nein

TLS_DHE_RSA_WITH_AES_128_GCM_SHA256	640	84,66%	ja	ja
TLS_RSA_WITH_AES_256_GCM_SHA384	632	83,60%	nein	nein
TLS_RSA_WITH_AES_128_GCM_SHA256	624	82,54%	nein	nein
TLS_DHE_RSA_WITH_AES_256_CBC_SHA256	623	82,41%	ja	nein
TLS_DHE_RSA_WITH_AES_256_CBC_SHA	621	82,14%	nein	nein
TLS_DHE_RSA_WITH_AES_128_CBC_SHA256	615	81,35%	ja	nein
TLS_DHE_RSA_WITH_AES_128_CBC_SHA	614	81,22%	nein	nein
TLS_DHE_RSA_WITH_AES_256_GCM_SHA384	600	79,37%	ja	ja
TLS_DHE_RSA_WITH_CAMELLIA_256_CBC_SHA	487	64,42%	nein	nein
TLS_DHE_RSA_WITH_CAMELLIA_128_CBC_SHA	487	64,42%	nein	nein
TLS_RSA_WITH_CAMELLIA_128_CBC_SHA	486	64,29%	nein	nein
TLS_RSA_WITH_CAMELLIA_256_CBC_SHA	481	63,62%	nein	nein
TLS_ECDHE_RSA_WITH_CHACHA20_POLY1305_SHA256	205	27,12%	nein	ja
TLS_CHACHA20_POLY1305_SHA256	203	26,85%	nein	ja
TLS_AES_256_GCM_SHA384	203	26,85%	nein	ja
TLS_AES_128_GCM_SHA256	203	26,85%	nein	ja
TLS_DHE_RSA_WITH_CHACHA20_POLY1305_SHA256	173	22,88%	nein	ja
TLS_RSA_WITH_3DES_EDE_CBC_SHA	171	22,62%	nein	nein
TLS_RSA_WITH_SEED_CBC_SHA	170	22,49%	nein	nein
TLS_ECDH_anon_WITH_AES_256_CBC_SHA	170	22,49%	nein	nein
TLS_ECDH_anon_WITH_AES_128_CBC_SHA	170	22,49%	nein	nein
TLS_RSA_WITH_AES_256_CCM	165	21,83%	nein	nein
TLS_RSA_WITH_AES_128_CCM	165	21,83%	nein	nein
TLS_DHE_RSA_WITH_SEED_CBC_SHA	160	21,16%	nein	nein
TLS_RSA_WITH_AES_256_CCM_8	152	20,11%	nein	nein
TLS_RSA_WITH_AES_128_CCM_8	152	20,11%	nein	nein
TLS_DHE_RSA_WITH_AES_256_CCM	145	19,18%	ja	ja
TLS_DHE_RSA_WITH_AES_128_CCM	145	19,18%	ja	ja
TLS_DHE_RSA_WITH_AES_256_CCM_8	136	17,99%	nein	nein

TLS_DHE_RSA_WITH_AES_128_CCM_8	136	17,99%	nein	nein
TLS_ECDHE_RSA_WITH_CAMELLIA_256_CBC_SHA384	135	17,86%	nein	nein
TLS_ECDHE_RSA_WITH_CAMELLIA_128_CBC_SHA256	135	17,86%	nein	nein
TLS_ECDHE_RSA_WITH_ARIA_256_GCM_SHA384	132	17,46%	nein	nein
TLS_ECDHE_RSA_WITH_ARIA_128_GCM_SHA256	132	17,46%	nein	nein
TLS_RSA_WITH_ARIA_256_GCM_SHA384	126	16,67%	nein	nein
TLS_RSA_WITH_ARIA_128_GCM_SHA256	126	16,67%	nein	nein
TLS_DHE_RSA_WITH_CAMELLIA_256_CBC_SHA256	126	16,67%	nein	nein
TLS_DHE_RSA_WITH_CAMELLIA_128_CBC_SHA256	126	16,67%	nein	nein
TLS_RSA_WITH_CAMELLIA_256_CBC_SHA256	119	15,74%	nein	nein
TLS_RSA_WITH_CAMELLIA_128_CBC_SHA256	119	15,74%	nein	nein
TLS_DHE_RSA_WITH_3DES_EDE_CBC_SHA	113	14,95%	nein	nein
TLS_DHE_RSA_WITH_ARIA_256_GCM_SHA384	110	14,55%	nein	nein
TLS_DHE_RSA_WITH_ARIA_128_GCM_SHA256	110	14,55%	nein	nein
TLS_ECDHE_RSA_WITH_3DES_EDE_CBC_SHA	100	13,23%	nein	nein
TLS_RSA_WITH_IDEA_CBC_SHA	73	9,66%	nein	nein
TLS_RSA_WITH_RC4_128_SHA	68	8,99%	nein	nein
TLS_RSA_WITH_RC4_128_MD5	58	7,67%	nein	nein
TLS_DH_anon_WITH_AES_256_CBC_SHA	57	7,54%	nein	nein
TLS_DH_anon_WITH_AES_128_CBC_SHA	57	7,54%	nein	nein
TLS_DH_anon_WITH_CAMELLIA_256_CBC_SHA	55	7,28%	nein	nein
TLS_DH_anon_WITH_CAMELLIA_128_CBC_SHA	55	7,28%	nein	nein
TLS_DH_anon_WITH_AES_256_GCM_SHA384	55	7,28%	nein	nein
TLS_DH_anon_WITH_AES_256_CBC_SHA256	55	7,28%	nein	nein
TLS_DH_anon_WITH_AES_128_GCM_SHA256	55	7,28%	nein	nein
TLS_DH_anon_WITH_AES_128_CBC_SHA256	55	7,28%	nein	nein
TLS_DH_anon_WITH_SEED_CBC_SHA	52	6,88%	nein	nein
TLS_ECDHE_RSA_WITH_RC4_128_SHA	46	6,08%	nein	nein
TLS_AES_128_CCM_SHA256	30	3,97%	nein	ja

TLS_DH_anon_WITH_3DES_EDE_CBC_SHA	24	3,17%	nein	nein
TLS_ECDH_anon_WITH_RC4_128_SHA	21	2,78%	nein	nein
TLS_ECDH_anon_WITH_3DES_EDE_CBC_SHA	21	2,78%	nein	nein
TLS_DH_anon_WITH_RC4_128_MD5	21	2,78%	nein	nein
TLS_AES_128_CCM_8_SHA256	20	2,65%	nein	nein
TLS_RSA_WITH_DES_CBC_SHA	5	0,66%	nein	nein
TLS_DHE_RSA_WITH_DES_CBC_SHA	5	0,66%	nein	nein
TLS_DH_anon_WITH_DES_CBC_SHA	4	0,53%	nein	nein
TLS_RSA_EXPORT_WITH_RC4_40_MD5	3	0,40%	nein	nein
TLS_RSA_EXPORT_WITH_RC2_CBC_40_MD5	3	0,40%	nein	nein
TLS_RSA_EXPORT_WITH_DES40_CBC_SHA	3	0,40%	nein	nein
TLS_DH_anon_EXPORT_WITH_RC4_40_MD5	3	0,40%	nein	nein
TLS_DH_anon_EXPORT_WITH_DES40_CBC_SHA	3	0,40%	nein	nein
TLS_DHE_RSA_EXPORT_WITH_DES40_CBC_SHA	3	0,40%	nein	nein
TLS_ECDHE_ECDSA_WITH_AES_256_GCM_SHA384	2	0,26%	ja	ja
TLS_ECDHE_ECDSA_WITH_AES_128_GCM_SHA256	2	0,26%	ja	ja
SSL_CK_RC4_128_WITH_MD5	1	0,13%	nein	nein
SSL_CK_DES_192_EDE3_CBC_WITH_MD5	1	0,13%	nein	nein

Tab. 4: Liste aller gesichteten Cipher Suiten

Beschreibung	BSI	IANA
TLS_RSA_WITH_AES_256_GCM_SHA384	nein	nein
TLS_RSA_WITH_AES_256_CBC_SHA256	nein	nein
TLS_RSA_WITH_AES_256_CBC_SHA	nein	nein
TLS_RSA_WITH_AES_128_GCM_SHA256	nein	nein
TLS_RSA_WITH_AES_128_CBC_SHA256	nein	nein
TLS_RSA_WITH_AES_128_CBC_SHA	nein	nein
TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384	ja	ja
TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384	ja	nein

TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA384	ja	nein
TLS_ECDHE_RSA_WITH_AES_256_CBC_SHA	nein	nein
TLS_ECDHE_RSA_WITH_AES_128_GCM_SHA256	ja	ja
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA256	ja	nein
TLS_ECDHE_RSA_WITH_AES_128_CBC_SHA	nein	nein

Tab. 5: Liste der Cipher Suiten von Outlook.com

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 639

Automatisierte Überprüfung von Webauftritten auf Fremdinhalte

Konstantin Knorr[1], David Müller[2]

Abstract: Die öffentlichen Webauftritte von Unternehmen und Behörden sind heute oft sehr umfangreich und entwickeln sich dynamisch weiter. In der dazugehörigen Datenschutzerklärung sollte der potentielle Besucher laut Artikel 13 der DS-GVO u.a. über die Einbindung externer Inhalte wie Fonts, Videos oder Werbung sowie über die dahinterstehenden Parteien informiert werden. Meist geschieht dies in Form einer einzigen Datenschutzerklärung für den gesamten Webauftritt. Insbesondere bei rasch wachsenden und sich dynamisch verändernden Webaufritten ist es schwierig, die Datenschutzerklärung aktuell zu halten. Es kommt leicht zu einer Über- oder Unterdeckung bei der Angabe der externen Parteien. Schlimmstenfalls drohen Bußgelder oder Abmahnungen bspw. durch Konkurrenten. Das Papier beschreibt ein Tool, das automatisiert in einer Tiefensuche die externen Verbindungen eines Webauftritts identifiziert und mit den manuell aus der Datenschutzerklärung ausgelesenen Parteien abgleicht. Nicht gelistete Fremdparteien verletzen die Informationspflicht und entbehren einer Rechtsgrundlage. Der ca. 32.000 Seiten umfassende Webauftritt einer Hochschule wurde mit dem Tool getestet. Die Ergebnisse des Tests werden beschrieben und diskutiert. Ferner werden Empfehlungen für das Vorgehen zur Beseitigung von Abweichungen gegeben. Die Autoren hoffen, damit den für die Datenschutzerklärung Verantwortlichen und Datenschutzbeauftragten eine Hilfestellung für die Erstellung und Pflege der Datenschutzerklärung ihres Webauftritts bieten zu können.

Keywords: Browser Automation, Datenschutzerklärung, HTTP, Python, Scrapy, Selenium, Web

1 Einleitung

Durch Aufruf einer Webseite in einem Browser werden personenbezogene Daten, wie zum Beispiel die IP-Adresse des verwendeten Rechners, an einen Server geschickt. Ist in dieser Webseite z.B. ein YouTube-Video eingebettet, so werden diese Daten außerdem an einen YouTube-Server gesendet. Verwendet die Webseite zusätzlich Google Fonts, werden die Daten auch noch an einen Google-Server weitergeleitet, vgl. Abbildung 1. Durch die Verwendung des World Wide Webs werden eine Vielzahl von Daten gesammelt und verarbeitet. Diese Daten können persistent gespeichert und unter anderem zum Web- Tracking [MM12] und zur Profilbildung genutzt werden. Die meisten Anwender sind sich dieser Verarbeitung ihrer Daten nicht bewusst. Individualisierte Werbung, die sich den Interessen des Nutzers anpasst, ist meist die einzige ersichtliche Auswirkung des Web-Trackings. Wie personenbezogene Daten genutzt werden dürfen, muss dem Nutzer zwar

[1] Hochschule Trier, Fachbereich Informatik, Am Schneidershof, D-54208 Trier, knorr@hochschule-trier.de
[2] Hochschule Trier, Fachbereich Informatik, Am Schneidershof, D-54208 Trier, muelled@hochschule-trier.de

doi:10.18420/inf2022_51

mitgeteilt werden, es schützt aber letztlich nicht vor Missbrauch.

Mayer und Mitchell geben einen Überblick der zum Web-Tracking eingesetzten Technologien [MM12]. Eine umfassende Studie mit mehr als einer Million untersuchten Seiten haben Englehardt und Narayanan 2016 präsentiert [EN16]. Ermakova et al. [EBFK18] analysieren in einer Literaturrecherche den aktuellen Forschungsstand bzgl. Web-Tracking. Wambach, Knorr und Schulte haben 2015 und 2016 die Start-Seiten von Hochschulen und Krankenhäusern auf externe Inhalte untersucht [WK15, WSK16]. Eine Diskussion der Ergebnisse im Vergleich zu dieser Studie findet sich in Abschnitt 5.

Der Beitrag beschreibt das Tool *wirechuck*, dass automatisiert die eingebundenen Drittparteien eines kompletten Web-Auftritts durch eine Tiefensuche ermittelt und gegen die in der Datenschutzerklärung (DSE) aufgelisteten Fremdparteien abgleicht. Die DSE eines Webauftritts kann somit geprüft und ggf. angepasst werden, was insbesondere für sich dynamisch weiterentwickelnde Webauftritte regelmäßig getan werden sollte. Eine DSE sollte (1) über die Fremdparteien informieren und (2) die Rechtsgrundlage dazu liefern. Bei einer nicht angegebenen Fremdpartei sind beide Punkte verletzt. Die Prüfung der Rechtsgrundlage verlangt juristische Expertise und erfolgt daher in der Regel manuell. Neben dem hier beschriebenen quell-offenen Ansatz zur Analyse von Webseiten und deren DSE gibt es kommerzielle Ansätze wie z.B. https://www.alfright.eu/, deren Funktionsweise allerdings nicht publiziert ist.

Der Rest des Artikels hat den folgenden Aufbau: Abschnitt 2 beschreibt die Grundlagen zu HTTP, URI und zum Datenschutz, insbesondere den Inhalt einer DSE von Webseiten. Das verwendete Tool *wirechuck* [wich] wird in Abschnitt 3 beschrieben. Die Studie des Webauftritts der Hochschule Trier folgt in Abschnitt 4. Abschließend diskutiert Abschnitt 5 die Ergebnisse, setzt sie in Bezug zu verwandten Arbeiten, spricht Empfehlungen aus und endet mit einem Ausblick bzgl. einer Weiterentwicklung des Tools.

2 Grundlagen

2.1 Technische Grundlagen: HTTP und URI

Das Hypertext Transfer Protocol (HTTP) ist ein vom World Wide Web Consortium standardisiertes Protokoll zur Übertragung von Daten auf der Anwendungsschicht. Das Protokoll wird vor allem zum Laden von Webseiten in einem Browser verwendet. HTTP setzt auf dem Transmission Control Protocol (TCP), neuerdings auch auf User Datagram Protocol (UDP) und dem Internet Protocol (IP) auf. Dies bedeutet, dass zwischen einem Client und einem Server eine Verbindung aufgebaut wird, über welche Daten ausgetauscht werden können. Die vom Client verwendete IP-Adresse gilt hierbei im Sinne des Datenschutzes als personenbezogenes Datum.

Beim Abruf einer Webseite in einem Browser ist im Body der Antwort vom Server in der

Regel eine HTML-Datei enthalten. Diese HTML-Datei beschreibt den Aufbau der Webseite und wird vom Browser interpretiert und dargestellt. Falls die Datei Bilder, Videos, Fonts oder sonstige Inhalte referenziert, müssen diese wiederum durch eine HTTP-Anfrage an denselben oder einen anderen Server nachgeladen werden. Dieser Vorgang ist in Abbildung 1 dargestellt. Bei populären Webseiten können schnell mehrere hundert Objekte von unterschiedlichen Servern geladen werden. Das vollständige Laden der Webseite dauert oft mehrere Sekunden.

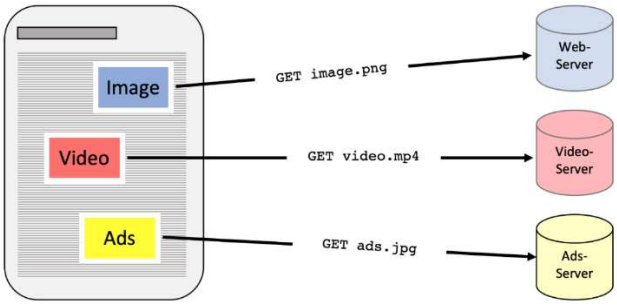


Abbildung 1: Einbindung externer Objekte auf einer Webseite.

Von besonderer Bedeutung beim Nachladen von Inhalten von Drittparteien ist hierbei der HTTP-Header. Dieser spezifiziert eine Anfrage oder Antwort und kann beispielsweise den Domain-Namen der Seite enthalten, die die benötigte Ressource referenziert hat (=Referer). Der Referer ist zwar eine freiwillige Angabe des Browsers, wird aber von den meisten Browsern in der Standard-Einstellung verwendet. Dies bedeutet, dass eine Drittpartei durch das Nachladen von Inhalten sowohl die IP-Adresse eines Clients über den IP-Header als auch die von ihm ursprünglich besuchte Webseite über den Referer im HTTP-Header erfahren kann. Parteien, die externe Objekte auf vielen Webseiten einbinden, erhalten so eine umfassende Übersicht der besuchten Webseiten des Clients. Es ist durch Einstellungen in der Datenschutz-Konfiguration des Browsers oder über Plug-Ins möglich, manche dieser Anfragen zum Nachladen von Inhalten zu unterdrücken. Firefox beispielsweise stellt dem Nutzer einen sog. Strict-Mode zur Verfügung, welcher bekannte Tracker blockiert und den Nutzer somit schützt.

Webseiten im Sinne dieses Artikels sind spezielle Uniform Resource Identifier (URI), die z.B. die folgende Form haben:

https://public.hochschule-trier.de/~knorr/AES/index.php?xy=fb

Dabei ist *https* (http über das kryptologische Protokoll TLS) das genutzte Protokoll. *public.hochschule-trier.de* ist die Domain, genauer eine Third Level Domain bzw. ein Rechnername in der Second Level Domain *hochschule-trier.de.* In der Studie in Abschnitt 4 wird auch der Begriff *Subdomain* verwendet. *~knorr/AES/* ist der Pfad zur Datei *index.php*. Hinter dem Fragezeichen wird an die Variable *xy* der Wert *fb* übergeben. Die Second Level Domain ist für die folgende Untersuchung fix. Jede Änderung in der Third Level Domain, im Pfad, in der referenzierten Datei und in den übergebenen Parametern

ergibt eine neue URI / Webseite, die eigene externe Objekte einbinden kann. Daher ist die Anzahl der untersuchten Webseiten schnell sehr groß.

2.2 Datenschutz

Unter dem Begriff Datenschutz wird der Schutz natürlicher Personen bei der Verarbeitung ihrer personenbezogenen Daten verstanden. Die Verarbeitung personenbezogener Daten ist in der Datenschutz-Grundverordnung, kurz DS-GVO, geregelt. Als personenbezogene Daten werden alle Informationen bezeichnet, die sich auf eine identifizierte oder identifizierbare Person beziehen (Artikel 4 DS-GVO). Bei Anwendbarkeit der DS-GVO sind Webseiten-Betreiber in Deutschland dazu verpflichtet, Angaben zu der Verarbeitung personenbezogener Daten und insbesondere Angaben zu der Weitergabe von personenbezogenen Daten an Dritte zu machen. Zu diesen personenbezogenen Daten zählt auch die IP-Adresse – alleine oder in Kombination mit dem Referer, da sich diese durch den entsprechenden Internetanbieter eindeutig einer Person zuordnen lässt (identifizierbare Person) [WSK16]. Im Allgemeinen dürfen nur Daten erhoben werden, die für festgelegte, eindeutige und legitime Zwecke genutzt werden. Diese Daten dürfen auch nur so lange gespeichert werden, wie es für diese Zwecke unbedingt notwendig ist (Artikel 5 DS-GVO).

Werden personenbezogene Daten verarbeitet, gilt nach Artikel 13 der DS-GVO eine Informationspflicht, der für Webseiten i.d.R. in Forum einer sog. DSE nachgekommen wird. Artikel 13 fordert u.a. Informationen über die folgenden Punkte:

- Namen und die Kontaktdaten des Verantwortlichen
- Kontaktdaten des Datenschutzbeauftragten
- Zwecke und Rechtsgrundlage der Verarbeitung
- Empfänger der Daten
- Übermittlung der Daten in ein Drittland oder an eine internationale Organisation
- Dauer der Speicherung
- Rechte der Betroffenen: Auskunft, Berichtigung, Löschung, Einschränkung, Widerspruch, Datenübertragbarkeit, Beschwerderecht, ggf. Widerrufsrecht

Bei der Einbindung von externen Objekten in Webseiten sind also die Empfänger (Firma und Firmensitz statt IP-Adresse) zu nennen. Übermittlungen in Drittländer müssen dabei gesondert berücksichtigt werden. Artikel 6 der DS-GVO listet die folgenden Möglichkeiten für die Rechtsgrundlage: Einwilligung, Erfüllung eines Vertrags, Erfüllung einer rechtlichen Verpflichtung, lebenswichtige Interessen, Wahrnehmung einer Aufgabe im öffentlichen Interesse und berechtigtes Interesse. Die Einwilligung als Rechtsgrundlage für Webseiten wird im §25 des neuen TTDSG (Gesetz über den Datenschutz und den Schutz der Privatsphäre in der Telekommunikation und bei Telemedien) explizit genannt. Eine Einwilligung muss der Verarbeitung vorausgehen. Sie wird i.d.R. über Checkboxen oder Zwei-Klick-Lösungen umgesetzt. Sie ist allerdings bzgl. der Verwaltung der Einwilligungen und möglichen Widerrufe aufwändig umzusetzen. Für Hochschulen mit Ihrem öffent-

lichen Bildungsauftrag kommt ebenso Artikel 6 Abs. 1 lit. c) („Wahrnehmung einer Aufgabe im öffentlichen Interesse“) in Betracht. Das „berechtigte Interesse“ ist laut DS-GVO für Behörden bzw. staatliche Hochschulen nicht anwendbar.

3 Das eingesetzte Tool *wirechuck*

Die Untersuchung eines Webauftritts auf externe Inhalte läuft in zwei Schritten ab. Der erste Schritt wird im Folgenden als *Crawling* (dt. „Kriechen / Krabbeln“) bezeichnet und dient der schnellen Identifizierung aller verlinkten Webseiten. Der zweite Schritt wird als *Scraping* (dt. „Kratzen“) bezeichnet und analysiert alle im Crawling identifizierten Seiten im Detail auf Fremdinhalte. Die Aufteilung in diese zwei Schritte erlaubt es, die Schritte unabhängig voneinander durchzuführen und unterschiedlich zu parametrisieren (z.B. für das Threading und die Delays). Diese Vorgehensweise erleichtert zudem die wiederholte Untersuchung eines Webauftritts, da die im ersten Schritt gefundenen Unterseiten wiederverwendet werden können. Für die Implementierung des Tools wurde die Programmier-Sprache Python verwendet.

Der Crawler erhält zwei Eingaben: (1) eine Start-URI und (2) eine zu durchsuchende Domain. Zu Beginn wird an die Start-URI eine HTTP-Anfrage gesendet. Die in der Antwort enthaltene HTML-Datei wird analysiert und auf Links (genauer `<a href> … </a>` im HTML-Code) zu weiteren Webseiten untersucht. Links auf Webseiten, die sich nicht in der angegebenen Domain befinden, werden ignoriert. An alle anderen Links wird wiederum eine HTTP-Anfrage gesendet. Dieser Vorgang wird solange wiederholt, bis keine weiteren Unterseiten mehr gefunden werden. Jedes Mal, wenn der Crawler eine HTTP-Antwort erhält, wird die in der dazugehörigen Anfrage enthaltene URI in eine Datei geschrieben (*<domain>.txt*). Wenn der gesamte Vorgang abgeschlossen ist, erhält der Nutzer als Ausgabe eine Text-Datei mit sämtlichen erreichbaren Unterseiten des durchsuchten Webauftritts. Für das Crawling wird das Python-Framework Scrapy (https://scrapy.org/) verwendet, welches für das Durchsuchen von Webseiten auf sogenannte Spider-Klassen (dt. „Spinne“) zurückgreift. Von besonderer Bedeutung für den Nutzer sind die Auszüge aus den Dateien settings.py und crawl_spider.py, die in Listing 1 und 2 abgebildet sind.

```
22. ROBOTSTXT_OBEY = True
…
30. DOWNLOAD_DELAY = 0.1
```

Listing 1: Auszug aus der Datei settings.py

```
11. rules = (
12.     Rule(LinkExtractor(allow=(), deny=('<Insert here>')),
    callback='parse_items', follow=True),
13. )
```

Listing 2: Auszug aus der Datei crawl_spider.py

ROBOTSTXT_OBEY legt fest, ob auf die robots.txt Datei des zu untersuchenden Webauftritts Rücksicht genommen werden soll und DOWNLOAD_DELAY gibt an, wie lange zwischen dem Herunterladen der einzelnen HTTP-Antworten gewartet werden soll. In Zeile 12 von Listing 2 kann ein Nutzer an Stelle von *<Insert here>* beliebige Strings angeben, die bei der Suche ignoriert werden sollen. Ein Beispiel hierfür wäre eine Subdomain, die nicht mit untersucht werden soll.

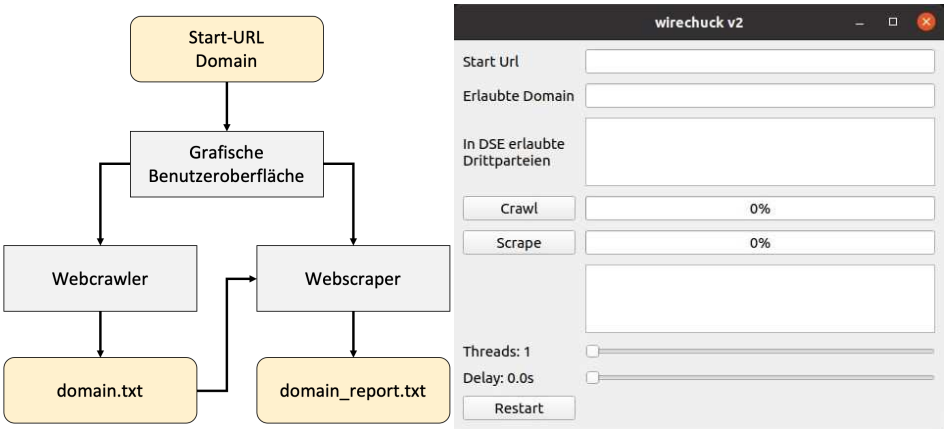


Abbildung 2: *wirechuck* Architektur (links) und Bedienoberfläche (rechts)

Der Scraper benötigt insgesamt vier Eingaben. (1) Die Domain wird verwendet, um den Namen der Datei *<domain>.txt* abzuleiten und die Datei auszulesen. Die in dieser Datei enthaltenen URIs werden anschließend der Reihe nach in einem Browser aufgerufen. Um den hierbei entstehenden Netzwerkverkehr aufzuzeichnen, wird ein Proxy-Server verwendet. Dies bedeutet, dass die HTTP-Anfragen vom Browser an diesen Server und anschließend vom Proxy-Server an die Ziel-Adresse gesendet werden. Jedes Mal, wenn eine dieser Webseiten vollständig geladen ist, werden alle entstandenen externen Verbindungen ausgelesen und gegen (2) eine Liste von erlaubten Drittparteien abgeglichen. Alle Verbindungen, die nicht explizit erlaubt sind, werden in eine Text-Datei mit dem Namen *<domain>_report.txt* geschrieben. Dieser Schritt dauert viel länger als das Crawling, da gewartet werden muss, bis alle Webseiten, inklusive externer Inhalte, vollständig geladen sind. Daher werden an dieser Stelle mehrere Threads verwendet, wobei in jedem Thread ein eigenes Browser-Fenster geöffnet wird. (3) Die Anzahl der zu verwendenden Threads sowie (4) die Dauer, die nach Öffnen einer Webseite gewartet werden soll (Delay), können vom Nutzer selbst festgelegt werden. Für die Implementierung des Scrapers wird das Framework Selenium (https://www.selenium.dev) verwendet, welches es dem Nutzer ermöglicht, Eingaben in einem Browser, in diesem Fall das Öffnen eines Browsers und der Aufruf einer bestimmten URI in diesem, zu automatisieren. Webdriver dafür werden von fast allen herkömmlichen Browsern angeboten, wobei hier Google Chrome verwendet wird.

Die Architektur und die Benutzeroberfläche des Tools sind in Abbildung 2 dargestellt.

Der Source Code und Installationshinweise stehen unter [wich] zur Verfügung. Müller [Mü22] beschreibt das Tool im Detail.

4 Studie an der Hochschule Trier

Die Hochschule Trier (HST) hat aktuell ~7.000 Studierende und ~750 Beschäftige. Sie hat mit Trier, Birkenfeld und Idar-Oberstein mehrere Standorte. Die HST nutzt zwei Second Level Domains: (1) hochschule-trier.de für die Standorte Trier und Idar-Oberstein und (2) umwelt-campus.de für den Standort Birkenfeld. Es gibt zwei Rechenzentren, eines in Trier, eines in Birkenfeld. Die Rechenzentren betreiben die Server für die Web-Infrastruktur selbst. Die DSE für die Webseiten der HST stammt vom Mai 2018, vgl. [DSEHST]. Darin werden keine externen Parteien als Empfänger von personenbezogenen Daten aufgeführt.

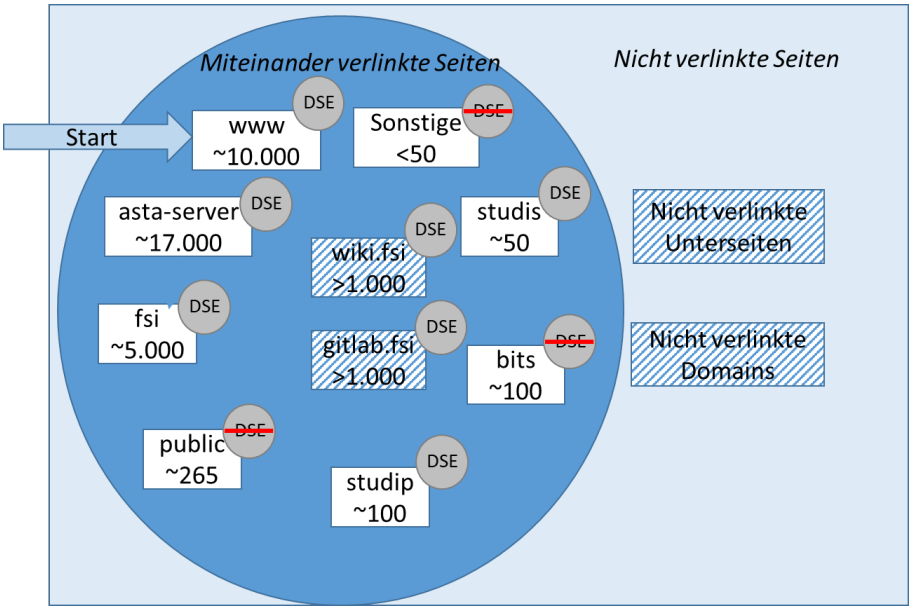


Abbildung 3: Öffentlich erreichbare Webseiten und deren Anzahl unter hochschule-trier.de. In der Studie exkludierte Unterseiten sind schraffiert dargestellt. Zusätzlich ist dargestellt, ob die Unterseiten eigene DSEn besitzen oder nicht.

Die Untersuchung des öffentlichen Teils des Web-Auftritts der Hochschule Trier wurde im Zeitraum 21.-24.02.2022 von einer externen IP-Adresse außerhalb des Hochschulnetzes durchgeführt. Als Start-URI wurde https://www.hochschule-trier.de und als Allowed-Domain hochschule-trier.de verwendet. Subdomains wie fsi.hochschule-trier.de und public.hochschule-trier.de wurden daher ebenfalls untersucht. Die Rechenzentren wurden vorab über die Untersuchung informiert. Für die Untersuchung des Webauftritts der Hochschule Trier wurde eine virtuelle Maschine mit 4GB RAM, 2 Prozessor-Kernen und einem

Ubuntu 20.04 Image verwendet.

4.1 Ergebnisse des Crawlings

Insgesamt wurden 32.574 Unterseiten in 2 Stunden und 26 Minuten gefundenen. Das entspricht ~220 gefundener Unterseiten pro Minute. Für das Crawling wurde auf die Vorgaben der *robots.txt* Datei Rücksicht genommen und es wurde ein Download-Delay von 0,1 Sekunden gewählt. Dies bedeutet, dass maximal 10 Anfragen pro Sekunde gesendet wurden. Abbildung 3 gibt einen Überblick der untersuchten Unterseiten. Alle Subdomains mit mehr als 50 Links sind aufgeführt.

4.2 Ergebnisse des Scrapings

Die Überprüfung der Unterseiten auf externe Verbindungen hat 32 Stunden und 2 Minuten gedauert. Hierfür wurden zwei Threads und ein Delay von 5 Sekunden verwendet. Pro Minute wurden also ~17 Unterseiten überprüft. Es wurden 69 Unterseiten mit externer Verbindung entdeckt. Die dazugehörigen externen Domains sind in Tabelle 1 dargestellt.

Domain	Anzahl Anfragen	Anzahl in Prozent
panopto-cache.zdv.net	1175	81,65%
www.youtube-nocookie.com	63	4,38%
www.gstatic.com	57	3,96%
fonts.gstatic.com	25	1,74%
www.youtube.com	18	1,25%
www.google.com	15	1,04%
f.vimeocdn.com	12	0,83%
i.vimeocdn.com	12	0,83%
i.ytimg.com	9	0,63%
yt3.ggpht.com	9	0,63%
fonts.googleapis.com	8	0,56%
fresnel.vimeocdn.com	8	0,56%
idp.fh-trier.de	7	0,49%
googleads.g.doubleclick.net	4	0,28%
player.vimeo.com	4	0,28%
static.doubleclick.net	2	0,14%
unpkg.com	2	0,14%
vimeo.com	2	0,14%
www.alfresco.com	2	0,14%
cdn.embed.ly	1	0,07%
cdnjs.cloudflare.com	1	0,07%
code.jquery.com	1	0,07%
ajax.googleapis.com	1	0,07%
dugie.de	1	0,07%

Tabelle 1: Externe Verbindungen nach Domain

5 Diskussion

Abbildung 3 zeigt eine Übersicht der über das Crawling erreichten Webseiten. *wirechuck* erkennt sehr zuverlässig alle verlinkten Seiten. Dies wurde bei der Entwicklung auch über einen eigens dafür erstellten Webauftritt geprüft. Das Tool kann allerdings nur die verlinkten Seiten finden. Nicht verlinkte Unterseiten oder nicht verlinkte Subdomains können nicht entdeckt werden. Subdomains kann es sehr viele geben. Für einen umfassenden Test müsste eine Liste mit allen Subdomains als Startadresse durchgetestet werden.

Es hat sich gezeigt, dass es einige Wikis und Webseiten zur Bereitstellung von Programm Code wie z.B. Gitlab gibt, die untereinander stark vernetzt sind. Hier genügt oft die Untersuchung der Startseite, da alle Seiten ähnlich aufgebaut sind. Bei den ersten Versuchen mit dem Tool stellte sich heraus, dass insbesondere die beiden Seiten fsi.hochschule-trier.de/wiki und gitlab.fsi.hochschule-trier.de mehr als 100.000 bzw. 60.000 Unterseiten besitzen. Aus Performanzgründen wurden diese beiden Seiten exkludiert. Während der Untersuchung wurden 17.000 Seiten unter asta-server.hochschule-trier.de untersucht. Auch hier ist ein enthaltenes Wiki die Ursache für die hohe Zahl an Seiten. Insgesamt sucht *wirechuck* also gründlich und teilweise zu sorgfältig, so dass einzelne Unterdomains ausgeschlossen werden müssen.

Bei der Untersuchung der Subdomains ist wichtig zu wissen, ob die Seiten eine eigene DSE aufweisen oder die Standard DSE der HST [DSEHST] verwenden. Das ist leider automatisiert noch nicht über das Tool feststellbar und muss manuell getan werden (vgl. Abbildung 3). Wird eine eigene DSE gefunden, muss diese gegen die entdeckten externen Verbindungen überprüft werden.

Beim Scraping zeigte sich, dass erfreulicherweise über 99% der untersuchten Webseiten keine externen Inhalte einbinden. Die DSE muss daher für diese Seiten nicht angepasst werden. Tabelle 1 zeigt die gefundenen externen Domains und die Häufigkeit ihres Aufrufs. panopto-cache.zdv.net ist die häufigste externe Adresse. Es handelt sich dabei um eine zentral für die Hochschulen in Rheinland-Pfalz vom Zentrum für Datenverarbeitung der Universität Mainz zur Verfügung gestellte Videoplattform. Diese wurde während der Pandemie-Jahre für die Aufzeichnung und Bereitstellung von Screencasts und Videos zu Lehrveranstaltungen intensiv genutzt. Mit der Universität Mainz besteht ein Vertrag zur Auftragsverarbeitung. Die nächsten fünf Domains sowie einige weitere Einträge wie i.ytimg.com und static.doubleclick.net kommen aus dem Google-Umfeld und sind im Hochschulkontext durchaus kritisch zu sehen [WK15]. Unter idp.fh-trier.de ist der Shibboleth-Identity-Provider zu finden, der vom Rechenzentrum der HST betrieben wird und aus technischen Gründen noch die alte Second Level Domain fh-trier.de statt hochschule-trier.de verwendet.

Etwa 10.000 der 32.000 Seiten werden über das freie Content-Management-System Typo3 (https://typo3.org/) erstellt. Dafür gibt es an der HST ein professionelles Team von Web-Redakteuren, die in der Vergangenheit hinsichtlich Datenschutz entsprechend sensibili-

siert und geschult wurden. Es gibt klare Verantwortlichkeiten und Prozesse für die Erstellung und Pflege der Typo3-Webseiten. Die meisten externen Objekte wurden auf Webseiten eingebunden, die nicht über Typo3 erstellt wurden. Das sind z.B. Webseiten für Forschungsprojekte, Konferenzen und Online-Berechnungstools, wie sie für Hochschulen und andere Forschungseinrichtungen typisch sind.

Die Überprüfung und ggf. die Beseitigung der eingebetteten Drittparteien ist nur eine Facette der zahlreichen Aufgaben, die aus Datenschutzsicht bzgl. eines Web-Auftritts anfallen. Der rasche Wechsel der eingesetzten Technologien im Webumfeld erschwert die Arbeit und macht eine kontinuierliche Überprüfung der Inhalte und Sensibilisierung der beteiligten Personen notwendig. Andere Themen im Webumfeld an der Hochschule in den letzten Jahren umfassen z.B. die analytische Auswertung der Webseitenbesuche, die Einbindung von Social Media in den Webauftritt und den Umgang mit Cookies.

5.1 Vergleich zu verwandten Arbeiten

Knorr, Wambach und Schulte haben in [WK15, WSK16] die Start-Webseiten von Krankenhäusern und Hochschulen untersucht. Damals vor der Anwendung der DS-GVO betteten 50% der Seiten externe Inhalte ein. 15% der untersuchten Webseiten hatten entweder keine oder eine fehlerhafte DSE. Bei den eingebetteten Drittparteien konnte eine starke Dominanz amerikanischer Unternehmen wie Google / Alphabet festgestellt werden. Dieser Trend wird auch von Englehardt und Narayanan in [EN16] bestätigt: “All of the top 5 third parties, as well as 12 of the top 20, are Google-owned domains. In fact, Google, Facebook, Twitter, and AdNexus are the only third-party entities present on more than 10% of sites.“ Auch diese Studie bestätigt diesen Trend. Im Unterschied zu den angeführten Publikationen ist die Anzahl der Fremdparteien einbettenden Seiten bei der Tiefensuche in unserer Fallstudie geringer. Dies liegt daran, dass in den Publikationen lediglich die Startseiten eines Webauftritts untersucht wurden und dass seit 2018 die DS-GVO angewendet wird.

5.2 Handlungsempfehlungen

Vorgelagert ist die Identifizierung der für die Webseiten verantwortlichen Personen. Da oft die Hochschulleitung im Impressum angegeben ist, kann über die Registrierungsdaten für die Subdomain, die beim Rechenzentrum hinterlegt sind, auf die Verantwortlichen geschlossen werden. Die Empfehlung an die Verantwortlichen ist: entweder (1) auf die Einbindung externer Objekte zu verzichten, (2) diese durch Alternativen zu ersetzen oder (3) die Erstellung einer eigenen DSE, die auf die externen Parteien hinweist.

Ad (1) und (2): Die Einbindung von externen Videos und Fonts wird dem Ersteller von Webseiten leichtgemacht. Meist geschieht dies in Anleitungen über ein Copy&Paste von entsprechendem HTML-Code. Auf die Datenschutzprobleme wird dabei oft unvollständig hingewiesen. Im Gespräch mit den Verantwortlichen stellte sich die fehlende Information

als häufigster Grund für die nicht dokumentierte Einbettung von externen Objekten heraus. Insbesondere die Verwendung externer Fonts kann in der Regel durch die lokale Bereitstellung von Fonts ersetzt werden. Für die Einbindung von Videos kann z.B. ein Bild mit hinterlegtem Link auf das externe Video (am besten mit Hinweis auf die externe Quelle) statt einer direkten Einbindung des Videos oder eine Zwei-Klick-Lösung, die die Einwilligung des Nutzers einholt, verwendet werden.

Ad (3): Die Erstellung einer eigenen DSE wird vom Datenschutzbeauftragten der Hochschule durch Text-Vorlagen und ein Schulungsvideo unterstützt. Die Erarbeitung einer eigenen DSE erhöht das Bewusstsein und das Verständnis für den Datenschutz und zeigt den Verantwortlichen zusätzlich das Spannungsfeld zwischen aktuellen Web-Technologien und dem Datenschutz deutlich auf.

5.3 Ausblick

Die automatische Identifikation von Fremdinhalten eines gesamten Web-Auftritts erleichtert die Prüfung der DSE für die verantwortlichen Parteien wie die Datenschutzbeauftragten enorm. Verstöße werden unmittelbar anzeigt. Eine regelmäßige Prüfung ist möglich. Das vorgestellte Tool kann eine Über- oder Unterdeckung zuverlässig feststellen. Für die Zukunft sind folgende Verbesserungen denkbar:

- Die Überprüfung des internen Webauftritts (Intranet), der oft nur nach vorheriger Authentisierung z.B. über Shibboleth erreichbar ist, wird noch nicht unterstützt. Die Herausforderungen liegt hierbei darin, die Anmeldung in Shibboleth zu automatisieren. Hierzu wird ein zusätzlicher Server, der sog. Identity Provider, genutzt. Ähnlich herausfordernd ist der Umgang mit dynamisch von Webseiten eingeblendeten Pop-Up-Fenstern, die z.B. Einwilligungen zu Cookies einfordern.
- Im Zuge einer regelmäßigen Überprüfung eines Web-Auftritts wäre die Verwendung einer Datenbank statt Text-Dateien zur Speicherung der Ergebnisse wünschenswert. So könnten einzelne Seiten gezielt nachgeprüft werden und eine zeitliche Entwicklung der gefunden und gelösten Probleme aufgezeigt werden.
- Der Abgleich zwischen den Objekten, die extern eingebettet sind, und der in der DSE gelisteten externen Partner geschieht im vorgestellten Tool manuell (über Angaben zulässiger externer Domains). Das Auffinden und automatisierte Auslesen und Auswerten einer DSE ist eine Herausforderung. Ein standardisiertes, maschinenlesbares Format für eine DSE wäre dafür hilfreich. Insbesondere müsste für jede Fremdpartei mindestens das Tripel (1) Name / Anschrift der Partei, (2) genutzte Domain / IP-Adressen und (3) Rechtsgrundlage der Einbettung erfasst werden. Mit dem Projekt „Platform for Privacy Preferences“ (https://www.w3.org/P3P/) gab es dazu einen Vorschlag, der aber kaum genutzt und nicht mehr weiterentwickelt wird.

Literatur

[DSEHST] Datenschutzerklärung des Web-Auftritts der Hochschule Trier, https://www.hochschule-trier.de/datenschutz

[EBFK18] Ermakova, E., Bender, B., Fabian, B., Klimek, K.: "Web Tracking – A Literature Review on the State of Research", Proceedings of the 51st Hawaii International Conference on System Sciences, 2018

[EN16] Englehardt, E., Narayanan, A.: "Online Tracking: A 1-million-site Measurement and Analysis", Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, pp. 1388–1401, https://doi.org/10.1145/2976749.2978313

[MM12] Mayer, J., Mitchell, J.: "Third-Party Web Tracking: Policy and Technology", 2012 IEEE Symposium on Security and Privacy, 2012, pp. 413-427, https://doi.org/10.1109/SP.2012.47

[Mü22] Müller, D.: „Automatisierte Überprüfung von Datenschutzerklärungen bezüglich Fremdinhalten", Bachelorarbeit, Hochschule Trier, März 2022

[wich] Source Code und Installationsanleitung zum Tool *wirechuck*: https://seafile.rlp.net/d/af0df4d9566a4b5caa59/

[WK15] Wambach, T., Knorr, K.: „Technische Prüfung der Datenschutzerklärungen auf deutschen Hochschulwebseiten", Information Security Day (ISD), Würzburg, 2015.

[WSK16] Wambach. T., Schulte, L., Knorr, K.: „Einbettung von Drittinhalten im Web", DuD - Datenschutz und Datensicherheit 40(8): 523-527 (2016)

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 651

Natural Language Processing (NLP) und der Datenschutz - Chancen und Risiken für den Schutz der Privatheit

Inna Vogel[1], Tahireh Setz[2], Jeong-Eun Choi[3], Prof. Dr. Martin Steinebach[4]

Abstract: Maschinelle Lernverfahren können sowohl Chancen als auch Risiken für die Privatheit von Daten bedeuten. Zum einen können durch Techniken des Natural Language Processings personenbezogene Daten anonymisiert werden und zum anderen können maschinelle Lernmodelle selbst hinsichtlich der Identifizierbarkeit der darin enthaltenen Daten zum Risiko für die Anonymität werden. In dieser Arbeit werden beide Aspekte, auch im Kontext von Angriffen auf die KI, diskutiert und Lösungsansätze besprochen. Sodann wird die datenschutzrechtliche Dimension von Angriffen auf die KI dargestellt und relevante Vorschriften des Entwurfs der Künstlichen-Intelligenz-Verordnung beleuchtet. Denn auch hier stehen sich der datenschutzrechtliche Grundsatz der Datenminimierung und das Interesse am Erhalt der Datenqualität gegenüber - ein scheinbar dilemmatisches Verhältnis.

Keywords: Textverarbeitung, Natural Language Processing (NLP), Datenschutz, Privatheit, Maschinelles Lernen.

1 Einleitung

Ob in der Textübersetzung, Bilderkennung oder bei der Prozessoptimierung in der Industrie: das maschinelle Lernen (ML) hat sich in vielen Bereichen bewährt. Der Siegeszug des ML kann mit der gestiegenen Rechenleistung sowie den frei verfügbaren Datenmengen begründet werden. Allerdings müssen die datenschutzrechtlichen Regelungen für die Verarbeitung der Daten beachtet werden. Die Daten für das Training von ML-Verfahren enthalten in der Regel personenbezogene Informationen (z.B. Namen, Zitate natürlicher Personen, Adressen), bei deren Verarbeitung die Anforderungen der Datenschutzgrundverordnung (DSGVO) einzuhalten sind. Werden Personenbezüge entfernt, ist der Verarbeiter – zumindest rechtlich – freier in den Nutzungsmöglichkeiten. Für die Entwicklung von ML-Verfahren gibt es aus technischer Sicht zwei Möglichkeiten: die Verschleierung oder Entfernung personenbezogener Daten. Diese technischen Vorgänge korrespondieren mit den rechtlichen Konzepten der Pseudonymisierung und Anonymisierung, wobei sie nicht gleichzusetzen sind (zu den Unterschieden s. Kap.2.1).

In dieser Arbeit werden verschiedene Techniken der maschinellen Textverarbeitung (engl. Natural Language Processing, kurz „NLP“) zum Schutz der Privatheit vorgestellt. Hierzu gehört die Entitätenerkennung, Koreferenzauflösung sowie die Schreibstilverschleierung. Allerdings können auch ML-Verfahren selbst hinsichtlich der darin enthaltenen Daten

[134] Fraunhofer SIT, Rheinstraße 75, Darmstadt, 64295, [Vorname].[Nachname]@sit.fraunhofer.de
[2] Universität Kassel, Wirtschaftswissenschaften, Henschelstraße 4, Kassel, 34127, t.setz@uni-kassel.de

doi:10.18420/inf2022_52

zum Risiko für die Anonymität werden. Welche Schutzmechanismen es gegen Angriffe auf ML-Modelle gibt, wird in Kapitel 2.3 diskutiert. Kapitel 3 widmet sich datenschutzrechtlichen Aspekten. Hier wird ein rechtlicher Bezug zu den vorgestellten NLP-Verfahren sowie den Angriffen auf ML-Systeme und Schutzmechanismen hiergegen hergestellt. Der Verlust der Nutzbarkeit der Daten für Forschung und Praxis steht den Anforderungen des Datenschutzes gegenüber. Dieses scheinbar dilemmatische Verhältnis wird zum Schluss der Arbeit erläutert und Lösungsansätze dafür vorgestellt.

2 NLP und der Datenschutz

Maschinelle Lernverfahren bieten sowohl Chancen als auch Risiken für die Privatsphäre. Zum einen können ML-Verfahren zur Anonymisierung personenbezogener Daten genutzt werden und zum anderen können diese selbst hinsichtlich der Anonymität der darin enthaltenen Daten zum Risiko für die Anonymität werden. Beide Aspekte und wie der Datenschutz schon während der Datensammlung berücksichtigt werden kann, werden nachfolgend vorgestellt.

2.1 Anonymisierung/Pseudonymisierung während der Datensammlung

Um Informationen aus verschiedenen Interneträumen zu erfassen, werden sogenannte „Crawler“ eingesetzt. Hierbei handelt es sich um Programme, die sich automatisiert über Verlinkungen (URLs) auf Webseiten durch das Internet bewegen, Inhalte erfassen und herunterladen. Die Inhalte können nicht nur HTML-Webseiten, sondern auch andere Datenformate wie z.B. Text-, Bild-, Audio- und Videodateien sein. Als „Scraper“ wird diejenige Teilkomponente eines Crawlers bezeichnet, welche die Analyse der erfassten Inhalte übernimmt. Einerseits müssen Verlinkungen für die nächsten Webseiten identifiziert und erfasst werden, andererseits muss eine Selektion der Inhalte auf Webseiten ermöglicht werden, um Daten zu erfassen, welche für das jeweilige Projekt relevant sind. Da Webseitenstrukturen individuell sind, wird für jede Webseite ein eigener Scraper benötigt, um die relevanten Inhalte adressieren zu können. Das bedeutet aber auch, dass bereits beim Crawlen die Verarbeitung personenbezogener Daten beschränkt und das Prinzip der Datenminimierung vorgenommen werden kann (Art. 5 Abs. 1 lit. c DSGVO). Zu den technisch-organisatorischen Maßnahmen der Datenminimierung kann die Pseudonymisierung gehören (Art. 25 Abs. 1 DSGVO). Demgegenüber ist auch die Anonymisierung stets zu prüfen. Anonymisierung meint die Veränderung personenbezogener Daten, sodass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßigen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können, siehe [RG21, S. 487]. Die in Art. 4 Nr. 5 DSGVO legal definierte Pseudonymisierung bewirkt die Funktionstrennung von Zuordnungs-information und Daten [Sc20, S. 284, 285].

Werden beispielsweise soziale Netzwerke wie Facebook gecrawlt, können personenbezogene Daten wie Profilname und -bild, Geburtsdatum, Wohnort, Namen der Follower und das Profilbild vom Crawler ignoriert werden, um den Datenschutz zu gewährleisten. Auf diese Weise kann die Weiterverarbeitung der Daten für Forschungszwecke ohne Personenbezug erfolgen (Art. 89 Abs. 1 Satz 4 DSGVO). Werden allerdings bestimmte Nutzerinformationen benötigt, um beispielsweise Korrelationen zwischen Usern zu erkennen, können personenbezogene Daten auch während des Crawlens pseudonymisiert werden und zusätzlich verschlüsselt gespeichert werden [St20]. Dennoch muss erwähnt werden, dass die Anonymisierung oder Pseudonymisierung von öffentlich zugänglichen Daten zwar eine risikominimierende Wirkung hat, die Zuordnung der betroffenen Person jedoch nicht ausgeschlossen werden kann. Um die Rückführung der Autorschaft zu minimieren, können stilverschleiernde Methoden in Betracht gezogen werden (siehe Kapitel 2.2).

2.2 Anonymisierung/Pseudonymisierung bei der Verarbeitung der erfassten Daten

Anders als bei strukturierten Daten (z.B. HTML-Code), wenn jedem Wert eine Kategorie zugeordnet werden kann, besitzen unstrukturierte Daten kein festes Schema oder kategorische Werte. Um schützenswerte Daten in unstrukturierten Freitextdokumenten (wie z.B. in Briefen, E-Mails, OP-Berichten oder Social Media-Kommentaren) zu anonymisieren/pseudonymisieren, können verschiedene Techniken des NLP angewandt werden. Um Entitäten wie Personen, Organisationen, Orte und numerische Daten (wie z.B. Geburtsdatum, Telefonnummer, Körpermaßeinheiten etc.) automatisiert zu identifizieren, kann die Eigennamenerkennung (engl. Named-Entity-Recognition, kurz „NER“) eingesetzt werden. Die besten Ergebnisse erzielt momentan „Flair“[5], ein Modell von Hugging Face, welches auf einem vortrainierten Transformer basiert, einer 2019 vorgestellten Deep-Learning-Architektur [SA20, Ak19]. Das Verfahren wurde auf den CoNLL-2003 Datensätzen in verschiedenen Sprachen trainiert [SA20]. Bei einer Vier-Klassen-Erkennung[6] werden F1-Werte für die deutsche Sprache von 92,31 erreicht, für die englische Sprache 94,36 (siehe Tab. 1).

	CoNLL-2003 (Deutsch) [SA20]	CoNLL-2003 (Eng.) [SA20]	Twitter (Deutsch) [St19]
FLAIR	92,31	94,36	82,01

Tab. 1: F1-Score von FLAIR [SA20], eigene Evaluierung auf Twitter-Posts [St19]

Die Genauigkeit bei der Erkennung von Personennamen sinkt, wenn das Verfahren auf Twittertexten angewandt wird. Eine manuelle Evaluierung auf 400 Twitter-Posts hat einen

[5] Flair-Modell von Hugging Face: https://huggingface.co/flair/ner-german-large
[6] NER Vier-Klassen: Person, Organisation, Ort und „Miscellaneous“ (alle weiteren Eigennamen)

F1-Wert von 82,01 ergeben (siehe Tab. 1) [St19]. Das liegt zum einen daran, dass bei Twitter weniger auf Rechtschreibung und Grammatik geachtet wird, aber auch daran, dass oft keine Standardsprache verwendet wird. Weiterhin gelten plattformspezifische Regeln wie die Verwendung von Hashtags oder @-Mentions, was eine Entitätenerkennung zusätzlich erschwert.

Einen höheren Datenschutz bietet die Koreferenzauflösung (engl. Coreference Resolution), welche zur Aufgabe hat, Relationen zwischen Textelementen zu finden, die auf dieselbe Entität verweisen. Beispielsweise verweist der Ausdruck „*die ehemalige Bundeskanzlerin*“ auf die Person „*Angela Merkel*“. Aber auch Pronomen (z.B. „*sie*“, „*seine*“) und andere Ausdrücke können referierend sein (z.B. verweist „*Onlineversand-händler*“ auf „*Amazon"*) [LHZ18]. Coreference Resolution kann eingesetzt werden, um einen höheren Datenschutz zu gewährleisten. So werden nicht nur Entitäten erkannt, sondern auch alle Begriffe, die sich auf die Entität beziehen, z.B. Berufsbezeichnungen und nähere Personenangaben wie „*die Krebspatientin*“, „*der Schlosser*“ oder „*der in Marburg lebende Jurastudent*“. Garat und Wonsever [GW22] haben die Koreferenzauflösung z.B. dazu eingesetzt, um spanische Gerichtsakten zu anonymisieren. Ziel war es nicht nur Personennamen, sondern auch alle Koreferenzen aufzulösen, um alle Erwähnungen derselben Person einheitlich zu pseudonymisieren. Die Schwierigkeit hierbei war ein Verfahren zu entwickeln, welches in der Lage ist unterschiedliche Schreibweisen von Personennamen zu erkennen. Hierfür wurde ein NER-Verfahren (F1-Score 90,21) zusammen mit einem Clusteringalgorithmus eingesetzt, um Koreferenzen aufzulösen (Accuracy 81). Allerdings merken die Autoren an, dass auf der Validierungsmenge die Genauigkeit bei der Erkennung und Verknüpfung der Namen mit ihren Koreferenzen immer noch unter 50% liegt. Generell lässt sich festhalten, dass Coreference Resolution-Verfahren niedrigere Genauigkeitswerte erzielen als NER-Verfahren. Das deutsche „E2E“[7]-Verfahren wurde auf verschiedenen Datensätzen (Literatur- und Nachrichtentexten) getestet und erzielt Genauigkeitswerte zwischen 64,72 (DROC-Datensatz) und 80,2 (CoNLL-Datensatz von 2012). Das CorefQA-System von Wu et al. [We20] ist derzeit das beste System für Englisch und erzielt auf dem CoNLL-2012 Datensatz einen F1-Score von 83,1 (siehe Tab. 2).

	TüBa (Deutsch)	SemEval (Deutsch)	DROC (Deutsch)	CoNLL-2012 (Eng.)
E2E (Deutsch)	78,79	74,46	64,72	80,2
CorefQA (Eng.)	-	-	-	83,1

Tab. 2: F1-Score für verschiedene Coreference Resolution-Verfahren [SHB21, We20]

Selbst wenn personenbezogene Daten wie Name und E-Mail-Adresse im Text anonymisiert werden, kann der Autor anhand seines Schreibstils mit sogenannten

[7] „Neural End-to-end Coreference Resolution for German in Different Domains“

Autorschafts-Attributionssystemen (engl. Authorship Attribution) identifiziert werden. Um die Anonymität des Verfassers dennoch zu wahren, können Verfahren eingesetzt werden, die den Schreibstil im Text verfälschen (engl. Authorship Obfuscation). Ziel der Autorenverschleierung ist es, einen gegebenen Text automatisch so zu paraphrasieren, dass dieser weder von Menschen noch von modernen Verfahren zur Überprüfung der Autorschaft (engl. Authorship Verification) dem ursprünglichen Autor zugeordnet werden kann [Be20]. Es zeigt sich allerdings, dass viele Verfahren den Text so verändern, dass entweder der Kontext (d.h. die Semantik) zu stark verändert wird oder der Text gar unlesbar für den Menschen wird [HPS17, MSS20]. Mahmood et al. [MSS20] haben beispielsweise ein Verfahren vorgestellt, welches mit einem durchschnittlichen F1-Wert von 87 erkennen kann, ob ein Text verändert wurde oder nicht. Sie geben an, dass bestehende Methoden zur Verschleierung der Autorschaft selbst stilistische Signaturen hinterlassen, die mit Sprachmodellen identifiziert werden können. Ihre Ergebnisse verdeutlichen die Notwendigkeit, Methoden zur Verschleierung der Autorschaft weiterzuentwickeln, welche die Identität eines Autors, der Anonymität anstrebt, besser schützen.

2.3 Risiken für die Privatheit von maschinellen Lern- und Sprachmodellen

Trainierte maschinelle Lernmodelle sowie das Training dieser Modelle auf Daten mit Personenbezug bergen immer Risiken für die Privatheit. Lange wurde die Ansicht vertreten, dass ML-Verfahren die Anonymisierung von Trainingsdaten garantieren, da die Datensätze, die für das Training der ML-Algorithmen verwendet werden, in der Regel disjunkt mit den Datensätzen sind, welche in der Anwendung eingesetzt werden. Nur vom Modell Rückschlüsse auf die Trainingsdaten zu ziehen, sollte somit nicht möglich sein, was einer Anonymisierung der Trainingsdaten gleichkäme [WBH19, BG21]. Attacken auf ML-Modelle wie die „Membership Inference“ verwenden Datenpunkte, um zu rekonstruieren, ob diese für das Training des Modells verwendet wurden. Bei „Model-Inversion-Attacken“ können Angreifer zum einen für einen bestimmten Datenpunkt personenbezogene Attribute rekonstruieren, zum anderen sogar ganze Teile der Trainingsdaten wiederherstellen. Solche Angriffe stellen eine Verletzung der Privatheit dar, vor allem, wenn es um besondere Kategorien der personenbezogenen Daten nach Art. 9 Abs. 1 DSGVO geht und eine Anonymisierung ausdrücklich erforderlich ist. Solche Informationen im Trainingsdatensatz können beispielsweise mittels „Distillation“ geschützt werden. Hierbei wird ein Lernmodell verwendet, um einen weiteren Datensatz zu klassifizieren. Dieser neu erstellte Datensatz wird im Anschluss dazu verwendet, um ein neues separates Modell zu trainieren. Hierfür wird das Wissen des ursprünglichen Modells übertragen [Pa18]. Es ist allerdings zu beachten, dass alle Maßnahmen zum Schutz der Privatheit mit einer Reduktion der Klassifikationsgenauigkeit einhergehen [BG21].

Sprachmodelle (engl. Language Models) wie GPT-3[8] von OpenAI oder PaLM[9] von Google sind darauf trainiert, die Wahrscheinlichkeit aufeinanderfolgender Wörter im Satz zu berechnen und auf diese Weise unter minimalen Vorgaben Texte zu ergänzen und selbst zu verfassen. Die Grundlage der Sprachmodelle sind Hunderte von Gigabyte an Textdaten aus dem Web wie etwa E-Books, Wikipedia, Social-Media-Plattformen oder Reddit. Zwar haben die Sprachmodelle zu bahnbrechenden Verbesserungen in vielen NLP-Anwendungen – wie etwa der Textgenerierung oder -übersetzung – geführt, können allerdings sensible Unternehmens- und Kundendaten (wie Namen, Telefonnummern, Adressen usw.) preisgeben. Carlini et al. [Ca21] führten einen Proof-of-Concept-Angriff zur Extraktion von Trainingsdaten auf GPT-2 durch und zeigten, dass es möglich ist, die genannten privaten Informationen aus dem Modell zu extrahieren. Um dieses Problem zu minimieren, wäre ein Training der Sprachmodelle auf anonymisierten oder unproblematischen Daten denkbar, was bei der benötigten Datenmenge jedoch schwierig umzusetzen ist. Eine weitere Möglichkeit wäre der Einsatz von „Differential Privacy". Hierbei handelt es sich um ein Konzept, Datenmengen zu analysieren, ohne die Privatsphäre eines einzelnen Benutzers zu verletzen, indem die Originaldaten mithilfe von Hash-Verfahren (hashing), Anreicherung mit Zufallswerten (noise injection) und der Verwendung von Teilmengen (subsampling) so verändert werden, dass sie keine Rückschlüsse auf personenbezogene Daten mehr zulassen [Dw06]. Das Framework DPText [Be19] soll beispielsweise das Trainieren von Textrepräsentation ermöglichen, welche keine Rückschlüsse auf private Informationen ermöglichen. Der Algorithmus ADePT [KGD21] paraphrasiert den gegebenen Text und garantiert auf diese Weise Different Privacy im neu generierten Datensatz. Habernal zeigte jedoch, dass weder DPText noch ADePT Differential Privacy garantieren können [Ha21, Ha22]. Weiter merkt er an, dass es bei einem strengeren Datenschutz durchaus zu einem bedeutenden Performanzverlust kommt und die Wahl differenzierter Datenschutzmaßnahmen nach Bedarf an Datenschutz und Performanz beurteilt werden sollte [SIH21]. Auch werden durch die auf dem Konzept der Differential Privacy beruhenden Techniken die Originaldaten als solche nicht verändert und gelten rechtlich weiterhin als personenbezogene Daten [SB21][10]. Zudem werden durch das Verzerren der Daten Veränderungen verursacht, die zu einer Minderung der Datenqualität führen können. Sinnvoll könnte eine kombinierte Differential-Privacy-Garantielösung sein, wonach eine datenschutzrechtliche Garantie eine Voraussetzung ist, das Modell ohne Rauschen zu nutzen. Ohne Garantie ist nur eine durch Differential Privacy verrauschte Version erhältlich. Eine datenschutzrechtlich vorteilhafte Variante der Differential Privacy stellt das Jittering dar. Hierdurch werden Daten lediglich punktuell verzerrt, was – im Gegensatz zur Generalisierung/Randomisierung – nur zu einer minimalen Verzerrung des Gesamtbildes führt. Dafür werden ebenjene Einzeldaten ausgewählt, deren Änderungen relevant für die Privatsphäre sind. Durch eine präzise und geschickte Auswahl lässt sich

[8] GPT-3: https://openai.com/blog/openai-api/

[9] Pathways Language Model (PaLM): https://ai.googleblog.com/2022/04/pathways-language-model-palm-scaling-to.html

[10] Bretthauer, in: Spiecker gen. Döhmann/Bretthauer, Dokumentation zum Datenschutz, G. G 2 G 2.1 G 2.1.10 Stellungnahme 05/2014 zu Anonymisierungstechniken (WP 216) 3 3.1. 3.1.3.

sowohl eine repräsentative Genauigkeit der Ergebnisse als auch der Schutz der Privatsphäre erhalten, siehe S. 329, 382 in [Pa20].

3 Datenschutzrecht

Die maschinelle Textverarbeitung personenbezogener Daten muss sich im Rahmen der DSGVO bewegen. Der in Art. 5 Abs. 1 lit. c DSGVO verankerte Grundsatz der Datenminimierung kann u.a. durch die technisch-organisatorische Maßnahme der Pseudonymisierung verwirklicht werden. Jedoch kann sowohl durch Anonymisierung als auch Pseudonymisierung ein erheblicher Verlust der Datenqualität und damit Nutzbarkeit in Forschung und Praxis entstehen.

Gegenüber der Pseudonymisierung ist die Anonymisierung, die nach EG Nr. 26 S. 5 DSGVO die Anwendbarkeit der DSGVO entfallen lässt, grundsätzlich vorrangig anzuwenden [SHS19, Art. 89 DSGVO, Rn 51 f.]. Die Pseudonymisierung findet an verschiedenen Stellen der DSGVO Ausdruck, z.B. Art. 25 Abs. 1 und Art. 32 Abs.1 a). Für die privilegierte Datenverarbeitung zu wissenschaftlichen Zwecken gilt nach Art. 89 Abs. 1 DSGVO, dass technisch-organisatorische Maßnahmen wie die Pseudonymisierung zu treffen sind. Für besondere Kategorien personenbezogener Daten i.S.d. Art. 9 Abs. 1 DSGVO sieht das deutsche Recht in § 27 Abs. 3 BDSG Anonymisierung unter bestimmten Voraussetzungen ausdrücklich vor.

Pseudonyme Daten können – je nachdem, ob eine Person über die Zuordnungsinformation verfügt oder nicht - anonym oder personenbezogen sein [Ro18, S. 243, 246]. Gerade die anonymisierende Pseudonymisierung, bei der sämtliche Bezugspunkte zur Person aus den Daten entfernt werden, kann die Datenqualität jedoch erheblich beeinträchtigen. Zum Ausgleich dieses Spannungsverhältnisses kann hingegen die risikomindernde Pseudonymisierung nach Art. 32 Abs. 1 lit. a DSGVO von Vorteil sein [Vgl. Ro18, S. 245 f.]. In der nach Art. 32 Abs. 1 lit. a, Abs. 2 DSGVO zu erfolgenden Abwägungsentscheidung steht das Interesse des Datenverarbeiters am Erhalt qualitativ ergiebiger Daten den Grundrechten der natürlichen Personen, die von der Verarbeitung betroffen sein können, gegenüber. Denn im Rahmen der Beurteilung der Risiken und Folgen sind nach Art. 32 Abs. 2 DSGVO auch die mit der Verarbeitung selbst verbundenen Risiken des Verlusts einzubeziehen [Sy18, Art. 32 Rn. 10]. Gem. Art. 32 Abs. 1 S. 1 DSGVO ist der Zweck der Verarbeitung in der Risikoabwägung zu beachten. Werden Daten zu Forschungszwecken verarbeitet, kann der Schutz der Datenqualität stärker gewichtet werden als bei rein wirtschaftlichen Zwecken. Zwar steht dem Verantwortlichen und dem Auftragsverarbeiter sein Ermessen bei der Auswahl der Maßnahmen zu (Ebd.). Zu berücksichtigen ist aber die gesetzlich vorgegebene Zielrichtung der Abwägung: Die Risikobeurteilung nach Art. 32 DSGVO muss zum Zweck der Sicherstellung des angemessenen Schutzniveaus aus der Sicht der betroffenen Personen und nicht des Datenverarbeiters erfolgen [Sy18, Art. 32 Rn. 10; SHS19, Art. 32 DSGVO, Rn. 28].

3.1 Angriffe auf maschinelle Lernverfahren

Besonders zeigt sich das Problem des Verlusts der Datenqualität durch Pseudonymisierung bei maschinellen Lernmodellen. Einerseits hängt der wissenschaftliche bzw. wirtschaftliche Nutzen der Sprachmodelle [AK20, S. 24, 25] von der Qualität und Quantität der Trainingsdaten ab [KB20, S.462]. Andererseits fordert der Grundsatz der Datensparsamkeit nach Art. 5 Abs. 1 lit. c DSGVO ein Lernen mit so wenigen personenbezogenen Daten wie möglich [NK20, Rn 46], sodass eine ursprüngliche Anonymisierung oder ein Training mit wenigen Daten vorzuziehen wäre. Das besondere Risiko der Re-Identifikation besteht dadurch, dass durch Angriffe auf Sprachmodelle – entgegen bisheriger Annahmen - Rückschlüsse auf personenbezogene Trainingsdaten möglich sind (s.o.). Bei der Bestimmung der Identifizierbarkeit personenbezogener Daten sind gem. EG 26 zur DSGVO alle Mittel zu berücksichtigen, die von dem Verantwortlichen oder einer anderen Person nach allgemeinem Ermessen wahrscheinlich genutzt werden. Dabei sind auch Faktoren wie die Kosten der Identifizierung und der dafür erforderliche Zeitaufwand und die zum Zeitpunkt der Verarbeitung verfügbare Technologie und technologische Entwicklungen zu berücksichtigen.

Die technologische Entwicklung ermöglicht Angriffe gegen Sprachmodelle, die personenbezogene Trainingsdaten rekonstruieren. Ein besonderes Risiko der Identifizierung geht außerdem von Model-Inversion- und Inference-Attacken aus (siehe Kapitel 2.3), die deutlich machen, dass eine automatische Anonymisierung der Trainingsdaten nicht garantiert werden kann. Ob noch von einer faktischen Anonymität, die die Anwendbarkeit der DSGVO ausschließt, ausgegangen werden kann, ist indes fraglich [KB20, S.466 m.w.N]. Insbesondere ist zu beachten, dass Angriffe einen erheblichen Aufwand bedeuten. Zweifelhaft ist also, ob wegen des Aufwands von einer hinreichenden Wahrscheinlichkeit ausgegangen werden kann, dass Angriffe tatsächlich durchgeführt werden (Ebd., S. 467). Auch ist das Risiko, dass Rückschlüsse auf natürliche Personen gezogen werden, auf Grund der für Angriffe vorausgesetzten gewissen Kenntnis der Modelle als gering einzuschätzen, [KB20, S. 467] sodass jedenfalls eine faktische Anonymisierung angenommen werden kann, sofern im Einzelfall keine anderen Gründe entgegengehalten werden können.

Risiken durch wirtschaftlich oder anders motivierte Angriffe sind auch im Rahmen der Abwägung nach Art. 32 Abs. 1 lit. a, Abs. 2 DSGVO (s.o.) zu berücksichtigen. Abs. 2 nennt explizit auch Risiken durch unrechtmäßige Ereignisse, worunter auch Angriffe zählen [SHS19, Art. 32 DSGVO Rn. 59 m.w.N].

Schließlich wird vorgeschlagen, bestimmte Ausprägungen der willentlichen De-anonymisierung zu sanktionieren, z.B. durch Einführung eines Straftatbestandes auf nationaler Ebene [RG21, S. 487, 490].

3.2 KI-VO-E

Eine Lösung für das Problem des Verlusts der Datennutzbarkeit wird in Art. 10 Abs. 1 des EU-Kommissionsentwurfs für eine Verordnung über die Künstliche Intelligenz (KI-VO-E) vorgeschlagen. Dieser sieht vor, dass Hochrisiko-KI-Systeme i.S.d. Art. 6 KI-VO-E mit solchen Daten entwickelt werden müssen, die den in den Qualitätskriterien aus Abs. 2 - 5 entsprechen. Diese richten sich auch auf die Vollständigkeit der Daten. Abs. 2 zählt exemplarisch Schritte der Entwicklung und des Betriebs von Hochriko-KI-Systemen auf, bei denen geeignete Governance- und Datenverwaltungsverfahren gelten sollen. Dies betrifft nach Abs. 2 g) auch die Ermittlung von Datenlücken oder Mängeln und Methoden zur Behebung. Nach Abs. 3 S. 1 müssen Trainings-, Validierungs- und Testdatensätze relevant, repräsentativ, fehlerfrei und vollständig sein. Durch diese Governance-maßnahmen soll die Gefahr der Diskriminierung, die sich aus Verzerrungen in KI-Systemen ergeben könnte, gebannt werden (Vgl. EG 44).

Dass der Verordnungsgeber sich über die Risiken der Re-Identifizierung von Personen durch Rückgriff auf die Trainingsdaten bewusst ist, wird in EG 51 zur KI-VO-E deutlich. Hier wird von der Widerstandsfähigkeit von KI-Systemen gegenüber Versuchen böswilliger Dritter gesprochen, die die Schwachstellen der Systeme ausnutzen, indem sie z.B. auf Trainingsdatensätze zurückgreifen oder trainierte Modelle nutzen. Hiergegen sollten die Anbieter von Hochrisiko-KI-Systemen geeignete Maßnahmen ergreifen (z.B. IKT-Infrastruktur). Trotz des legitimen Zwecks sicherzustellen, dass Hochrisiko-KI-Systeme bestimmungsgemäß und sicher funktionieren und nicht zur Ursache für Diskriminierung werden (vgl. EG 44), würde der Erlass der vorgeschlagenen Vorschrift das Spannungsfeld zwischen Informationsgehalt und Datenschutz noch verstärken und zu einem dilemmatischen Verhältnis zwischen Erhalt der Datenqualität und dem Schutz personenbezogener Daten führen. Zum einen ist unklar, was unter „fehlerfrei“ zu verstehen ist [BM21, S. 276, 280; [RW21, S. 844, 847]. Die Gefahr besteht, dass ein während des Trainings hinzugefügtes Rauschen durch die zum Schutz personenbezogener Daten eingesetzte Differential Privacy (s. Kapitel 2.3) als Fehler gewertet werden könnte [Eb21, S. 528, 533], obwohl diese Maßnahme zum Schutz personenbezogener Daten sinnvoll wäre. Das Problem zeigt sich weiterhin in Abs. 5, wonach die Verarbeitung besonderer Kategorien personenbezogener Daten nach Art. 9 Abs. 1 DSGVO erlaubt ist, soweit es für die Beobachtung, Erkennung oder Korrektur von Verzerrungen in Hochrisiko-KI-Systemen unbedingt erforderlich ist (1.HS). Im 2. HS wird gefordert, dass angemessene Vorkehrungen für den Schutz der Grundrechte und Grundfreiheiten natürlicher Personen getroffen werden müssen, wozu auch die Pseudonymisierung gehört, wenn der verfolgte Zweck durch eine Anonymisierung erheblich beeinträchtigt würde. Offen bleibt, wie dieses Ziel zugleich mit der in Art. 10 Abs. 1 geforderten Vollständigkeit und Fehlerlosigkeit der Daten gewährleistet werden soll.

Bei Verstößen gegen Art. 10 KI-VO-E könnten nach Art. 71 III KI-VO-E Bußgelder bis zu 30.000.000 Euro oder 6% des weltweiten Jahresumsatzes verhangen werden. Bußgelder im Falle der willentlichen De-anonymisierung durch Angriffe gegen Modelle sind jedoch nicht vorgesehen.

4 Diskussion

Wir haben unterschiedliche Methoden der Datenanonymisierung mittels NLP-Methoden sowie während des Datensammelns (Crawling) vorgestellt. Wird die Koreferenzauflösung für die Anonymiserung von Daten verwendet, bei der sämtliche Bezugspunkte zur Person aus den Daten entfernt werden, kann die Datenqualität erheblich beeinträchtigt werden. NER könnte eine Lösung sein, welche mit weniger Informationsverlust einhergeht. Wenn auch der Textinhalt besonders schützenswert ist, können Techniken der Stilverschleierung eine Option sein. Dennoch bergen alle maschinellen Lernverfahren Gefahren, die beim Einsatz berücksichtigt werden sollten. Es existieren bislang keine ML-Verfahren, die eine hundertprozentige Zuverlässigkeit garantieren.

Zudem können ML-Verfahren selbst hinsichtlich der Anonymität der darin enthaltenen Daten zum Risiko werden, beispielsweise durch Membership-Inference oder Model-Inversion-Attacken. Um sich vor solchen Angriffen zu schützen, haben wir einige Ansätze wie Distillation, Differential Privacy und Jittering vorgestellt. Allerdings gehen diese Verfahren immer mit Performanz- und Datenqualitätsverlust einher. Und ob von einer faktischen Anonymität durch den Einsatz solcher Verfahren, die die Anwendbarkeit der DSGVO ausschließt, ausgegangen werden kann, ist indes fraglich. Andererseits ist das Risiko des Zugriffs auf personenbezogene Daten durch Angriffe auf ML-Systeme nach jetzigem Kenntnisstand als eher gering einzustufen. Denn solche Angriffe erfordern ein erhebliches technisches Wissen sowie eine begründete Motivation des Angriffs.

Der EU-Kommissionsentwurfs KI-VO-E sieht diesbezüglich vor, dass für Hochrisiko-KI-Systeme Datensätze entwickelt werden müssen, welche die folgenden Qualitätskriterien erfüllen: Vollständigkeit, Relevanz, Repräsentativität. Datenlücken und -mängel müssen geschlossen werden, wobei gleichzeitig von „*Widerstandsfähigkeit von KI-Systemen gegenüber Versuchen böswilliger Dritter*" gesprochen wird, was das Spannungsfeld zwischen Informationsgehalt und Datenschutz verstärkt. Zwar wird im KI-VO-E der Schutz besonderer Kategorien personenbezogener Daten und Systeme gefordert, was durch Einsatz von Differential Privacy ermöglicht werden könnte. Der Verordnungsentwurf lässt jedoch offen, wie zugleich eine Fehlerlosigkeit der Daten gewährleistet werden kann. Eine logische Schlussfolgerung wären NLP-Verfahren, welche nicht nur eine zuverlässige Datenanonymisierung bzw. -pseudonymisierung garantieren, sondern ebenfalls die Aufrechterhaltung der Datenqualität und Semantik. Auch der Schutz der Daten durch Jittering scheint eine minimalinvasive Methode zu sein, die sowohl eine hinreichende Datenqualität als auch Datenschutz gewährleistet [Pa20, S.376, 382].

5 Fazit

Die Entwicklung des maschinellen Lernens zeigt immer neue Risiken der Re-Identifizierung personenbezogener Daten und damit technische Grenzen der

Anonymisierung bzw. Pseudonymisierung auf. Gleichzeitig bringt der Einsatz stark anonymisierender bzw. pseudonymisierender Methoden die Gefahr des Verlusts der Nutzbarkeit der Daten für Forschung und Praxis mit sich. Hierdurch besteht ein Spannungsfeld zwischen dem datenschutzrechtlichen Prinzip der Datenminimierung, das den durch die Datenverarbeitung Betroffenen schützen soll, und dem Interesse des Verarbeiters an der Minimierung des Qualitätsverlusts. Dabei handelt es sich indes nur um ein scheinbar dilemmatisches Verhältnis: Gerade um die Sicherheit des Betroffenen zu gewährleisten (auch was den Schutz vor Diskriminierung durch Verzerrung von KI-Systeme angeht, vgl. EG 44 KI-VO-E), müssen der wissenschaftlichen Erforschung und anwendungsorientierten Entwicklung maschineller Lernmodelle hinreichend qualitative Daten zur Verfügung stehen. Dies wurde zwar auch in der KI-VO-E erkannt, dennoch müssen die Vorgaben des Art. 10 Abs. 1 KI-VO-E hinsichtlich der Datenqualität weniger streng sein, um einen ausreichenden Datenschutz zu gewährleisten. Dies gilt gerade auch dafür, um bessere datenschutzfreundliche technische Maßnahmen entwickeln zu können und andere grundrechtlich gebotenen Ziele (z.B. Vermeidung von Diskriminierung) zu erreichen.

Danksagung

Diese Forschungsarbeit wurde vom Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Projekts DYNAMO unterstützt.

Literatur

[Ak19] Akbik, A.; Bergmann, T.; Blythe, D.; Rasul, K.; Schweter, S.; Vollgraf, R.: FLAIR: An easy-to-use framework for state-of-the-art NLP. In: NAACL 2019, 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations). S. 54-59, 2019.

[AK20] Apel, S.; Kaulartz, M.: Rechtlicher Schutz von Machine Learning-Modellen. RDi 1/1, S. 24-34, 2020.

[Be19] Beigi, G.; Shu, K.; Guo, R.; Wang, S.; Liu, H.: Privacy Preserving Text Representation Learning. In: Proceedings of the 30th ACM Conference on Hypertext and Social Media. HT '19, Association for Computing Machinery, Hof, Germany, S. 275-276, 2019, url: https://doi.org/10.1145/3342220.3344925.

[Be20] Bevendorff, J.; Wenzel, T.; Potthast, M.; Hagen, M.; Stein, B.: On divergence-based author obfuscation: An attack on the state of the art in statistical authorship verification, it - Information Technology 62/2, S. 448-452, Juli 2021.

[BG21] Battis, V.; Graner, L.: Risiken für die Privatheit aufgrund von Maschinellem Lernen. In (Reussner, R. H.; Koziolek, A.; Heinrich, R., Hrsg.): INFORMATIK 2020. Gesellschaft für Informatik, Bonn, S. 841-855, 2021.

[BM21] Bomhard, D.; Merkle, M.: Europäische KI-Verordnung. Der aktuelle Kommissionsentwurf und praktische Auswirkungen, RDi 2/6, S. 276-283, 2021.

[Bo21] Boenisch, F.: Privatsphäre und Maschinelles Lernen. Über Gefahren und Schutzmaßnahmen, Datenschutz und Datensicherheit - DuD 45/7, S. 448-452, Juli 2021.

[Ca21] Carlini, N.; Tramèr, F.; Wallace, E.; Jagielski, M.; Herbert-Voss, A.; Lee, K.; Roberts, A.; Brown, T.B.; Song, D.X.; Erlingsson, Ú.; Oprea, A.; Raffel, C.: Extracting Training Data from Large Language Models. USENIX Security Symposium, 2021.

[Ch20] Chai, H.; Zhao, W.; Eger, S.; Strube, M.: Evaluation of Coreference Resolution Systems Under Adversarial Attacks. In: Proceedings of the First Workshop on Computational Approaches to Discourse. Association for Computational Linguistics, Online, S. 154-159, Nov. 2020, url: https://aclanthology.org/2020.codi-1.16.

[Dw06] Dwork, C.: Differential Privacy. In (Bugliesi, M.; Preneel, B.; Sassone, V.; Wegener, I., Hrsg.): Automata, Languages and Programming. Springer Berlin Heidelberg, Berlin, S. 1-12, 2006.

[Eb21] Ebers, M.; Hoch, V.; Rosenkranz, F.; Ruschemeier, H.; Steinrötter, B.: Der Entwurf für eine EU-KI-Verordnung: Richtige Richtung mit Optimierungsbedarf. Eine kritische Bewertung durch Mitglieder der Robotics & AI Law Society (RAILS), RDi 2/11, S.528-537, 2021.

[ES21] Ebert, A.; Spieker gen. Döhmann, I.: Der Kommissionsentwurf für eine KI-Verordnung der EU. Die EU als Trendsetter weltweiter KI-Regulierung, NVwZ 40/16, S. 1188-1193, 2021.

[GW22] Garat, D.; Wonsever, D.: Automatic Curation of Court Documents: Anonymizing Personal Data. Information 13/27, 2022, url: https://doi.org/10.3390/info13010027.

[Ha21] Habernal, I.: When differential privacy meets NLP: The devil is in the detail. In: Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Online und Punta Cana, S. 1522-1528, Nov. 2021, url: https://aclanthology.org/2021.emnlp-main.114.

[Ha22] Habernal, I.: How reparametrization trick broke differentially-private text representation learning. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Dublin, (to appear), 2022, url: https://arxiv.org/abs/2202.12138.

[HPS17] Hagen, M.;Potthast, M.; Stein B.:Overview of the Author Obfuscation Task at PAN 2017: Safety Evaluation Revisited. In (Cappellato, L.; Ferro, N.; Goeuriot, L.; Mandl, T., Hrsg.): Working Notes of CLEF 2017 – Conference and Labs of the Evaluation Forum, Dublin, September 11-14, 2017. Bd. 1866. CEUR Workshop Proceedings, CEUR-WS.org, 2017.

[KB20] Kaulartz, M.; Braegelmann, T.: Rechtshandbuch Artificial Intelligence und Machine Learning. Beck C. H., 2020.

[KGD21] Krishna, S.; Gupta, R.; Dupuy, C.: ADePT: Auto-encoder based Differentially Private Text Transformation. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. Association for Computational Linguistics, Online, S. 2435-2439, Apr. 2021, url:

https://aclanthology.org/2021.eacl-main.207.

[LHZ18] Lee, K.; He, L.; Zettlemoyer, L.: Higher-Order Coreference Resolution with Coarse-to-Fine Inference. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers). Association for Computational Linguistics, S. 687-692, 2018.

[MSS20] Mahmood, A.; Shafiq, Z.; Srinivasan, P.: A Girl Has A Name: Detecting Authorship Obfuscation. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, S. 2235-2245, Juli 2020, url: https://aclanthology.org/2020.acl-main.203.

[NK20] Niemann, F.; Kevekordes, J.: Machine Learning und Datenschutz (Teil 1): Grundsätzliche datenschutzrechtliche Zulässigkeit. Computer und Recht 36/1, S. 17-25, 2020, url: https://doi.org/10.9785/cr-2020-360110.

[Pa18] Papernot, N.; Song, S.; Mironov, I.; Raghunathan, A.; Talwar, K.; Erlings-son, Ú.: Scalable Private Learning with PATE. In: 6th International Conference on Learning Representations, ICLR 2018, Vancouver, April 30 - May 3, 2018, Conference Track Proceedings, 2018.

[Pa20] Papastefanou, S.: „Database Reconstruction Theorem“ und die Verletzung der Privatsphäre (Differential Privacy). Computer und Recht 36/6, S. 379-386, 2020, url: https://doi.org/10.9785/cr-2020-360610.

[RG21] Roßnagel, A.; Geminn, C.: Vertrauen in Anonymisierung. Regulierung der Anonymisierung zur Förderung Künstlicher Intelligenz, Zeitschrift für Datenschutz - ZD 11/9, S. 487-490, 2021.

[Ro18] Roßnagel, A.: Pseudonymisierung personenbezogener Daten, ZD 9/6, S.241-288, 2018.

[RW21] Roos, P.; Weitz, C. A.: Hochrisiko-KI-Systeme im Kommissionsentwurf für eine KI-Verordnung. IT- und produktsicherheitsrechtliche Pflichten von Anbietern, Einführern, Händlern und Nutzern, MMR 24/11, S. 844-850, 2021.

[SA20] Schweter, S.; Akbik, A.: FLERT: Document-Level Features for Named Entity Recognition, 2020, arXiv: 2011.06993 [cs.CL].

[SB21] Spiecker gen. Döhmann, I.; Bretthauer, S., Hrsg.: Dokumentation zum Datenschutz, 84.Aufl. Baden-Baden 2021.

[Sc20] Schleipfer, S.: Pseudonymität in verschiedenen Ausprägungen. Wie gut ist die Unterstützung der DS-GVO? ZD 10/6, S. 284-290, 2020.

[SHB21] Schröder, F.; Hatzel, H. O.; Biemann, C.: Neural End-to-end Coreference Resolution for German in Different Domains. In Proceedings of the 17th Conference on Natural Language Processing (KONVENS 2021), S. 170-181, Düsseldorf, 2021.

[SHS19] Simitis, S.; Hornung, G.; Spiecker Döhmann, I., Hrsg.: Datenschutzrecht, DSGVO mit BDSG. Nomos, Baden-Baden, 2019.

[SIH21] Senge, M.; Igamberdiev, T.; Habernal, I.: One size does not fit all: Investigating strategies for differentially-private learning across NLP tasks. CoRRabs/2112.08159/, 2021, arXiv: 2112.08159, url: https://arxiv.org/abs/2112.08159. [St19] Struß, J. M.; Siegel, M.; Ruppenhofer, J.; Wiegand, M.; Klenner, M.: Overview of GermEval

Task 2, 2019 Shared Task on the Identification of Offensive Language. In: Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019). German Society for Computational Linguistics & Language Technology, Erlangen, S. 354-365, 2019.

[St20] Steinebach, M. et al.: Kapitel 4: Automatisierte Erkennung von Desinformationen. In: Steinebach, M.; Rinsdorf, L.; Krämer, N.; Roßnagel, A.: Desinformation aufdecken und bekämpfen: Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität. Baden-Baden: Nomos Verlagsgesellschaft (Schriften zum Medien- und Informationsrecht 45). S. 101-148, 2020.

[Sy18] Sydow, G.: Hrsg.: Europäische Datenschutzgrundverordnung, Handkommentar, 2. Aufl., Baden-Baden: Nomos Verlagsgesellschaft, 2018.

[WBH19] Winter, C.; Battis, V.; Halvani, O.: Herausforderungen für die Anonymisierung von Daten. In (David, K. et al., Hrsg.). INFORMATIK 2019. Gesellschaft für Informatik e.V., Bonn, S. 9-52, 2019.

[We20] Wu, W.; Wang, F.; Yuan, A.; Wu, F.; Li, J.: CorefQA: Coreference resolution asquery-based span prediction. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, S. 6953-6963, 2020.

Deepfakes im VideoIdent-Verfahren: (fehlende) Straf- und zivilrechtliche Konsequenzen für Täter

Simone Salemi[1], Bianca Steffes[2]

Abstract: Der Einsatz des sogenannten VideoIdent-Verfahrens zur Authentifzierung im Videochat erfreut sich wachsender Beliebtheit bei Banken und Versicherungen. Nach erfolgter Legitimation wird der Zugang zu neu eröffneten Bankkonten freigeschaltet. Gleichzeitig führen Fortschritte im Bereich des Deep Learnings dazu, dass Manipulationen von Videos mittels sogenannter Deepfakes kaum mehr erkennbar sind. Dieser Beitrag widmet sich der Frage, ob Deepfakes eine reale Gefahr für die Sicherheit des VideoIdent-Verfahrens darstellen und wie der Einsatz rechtlich zu bewerten ist.[3]

Keywords: Deepfakes; Authentifizierung; VideoIdent

1 Problemstellung

Die Manipulation von Videos stellt mit heute gängigen technischen Mitteln selbst für Laien keine große Herausforderung mehr dar. Mittels sogenannter *Deepfakes* können täuschend echte Fälschungen erstellt werden, die mit dem bloßen Auge kaum mehr erkennbar sind. Dem Einsatz von Deepfakes sind indes nur wenige Grenzen gesetzt. Zu denken wäre an den Einsatz im VideoIdent-Verfahren, welches inzwischen von mehreren Banken (vgl. [Co], [IN]) zur Authentifizierung mittels Videochat genutzt wird. Nach erfolgter Legitimation kann bspw. ein Girokonto eröffnet werden. Betrachtet man dieses Verfahren nun unter Berücksichtigung der Manipulationsmöglichkeiten durch den Deepfake-Einsatz, erscheint es alles andere als sicher. Dass sich eine Person etwa der Identität einer anderen bedient, um ein Konto auf deren Namen zu eröffnen, ist durchaus vorstellbar. Überzieht diese Person nun besagtes Konto, bleibt der Bank als Ansprechpartner nur die Person, als die sich der Täter ausgegeben hat. In diesem beispielhaften Szenario werden also gleich mehrere Beteiligte geschädigt, da sowohl die Person, deren Identität sich der Täter angeeignet hat, als auch die Bank betroffen sind. Aus technischer Sicht ist zu klären, ob der Deepfake-Einsatz im VideoIdent-Verfahren realistisch umsetzbar ist. Rechtliche Regelungen, die speziell auf den Einsatz von Deepfakes zugeschnitten sind, existieren derweil nicht. Zu prüfen ist daher, inwieweit die existierenden Regelungen sich auf den beschriebenen Fall anwenden lassen.

[1] Saarbrücker Zentrum für Recht und Digitalisierung, Campus C3.1, 66123 Saarbrücken, Deutschland, simone.salemi@zrd-saar.de

[2] Universität des Saarlandes, Lehrstuhl für Rechtsinformatik, Saarland Informatics Campus, 66123 Saarbrücken, Deutschland, bianca.steffes@uni-saarland.de

[3] Dieser Beitrag wurde durch das Projekt „Digitale Präsenz bei Gericht“ (BMJ) inspiriert.

doi:10.18420/inf2022_53

2 Technische Grundlagen

In einem ersten Schritt soll nun erläutert werden, was genau Deepfakes aus technischer Sicht sind und auf welchen technischen Methoden sie basieren. Im weiteren Verlauf soll bewertet werden, ob das einleitend beschriebene Szenario in Betracht des heutigen Stands der Technik als realistisch einzustufen ist. Der Ursprung des Begriffs „Deepfake“ ist im Jahr 2017 zu verordnen, in dem auf der Plattform reddit.com ein Nutzer mit dem Namen „deepfakes“ begann, mehrere Deepfake-Videos zu posten (bspw. bei [Sa17]). In diesen Videos wurden die Gesichter der Darsteller durch die von bekannten Persönlichkeiten ausgetauscht. Dies beschreibt eine der möglichen Ausprägungen von Deepfakes: das *Face Swapping*. Neben dem reinen Austauschen eines Gesichts in einem Bild oder Video können Deepfakes aber auch Sprachartefakte sein, die eine Person nie gesprochen hat, oder auch komplett neue Videos, in denen Mimik und Kopfbewegungen von Personen vollkommen nach Belieben gesteuert werden können (*Face Reenactment*). Zhang [Zh22] fasst die Menge an möglichen Deepfakes aus dem Englischen übersetzt als „realistisch wirkende aber unechte Bild-, Audio-, Video- und weitere digitale Medien, die durch KI-Methoden und insbesondere deep learning erstellt wurden“ zusammen. Das Bundesamt für Sicherheit in der Informationstechnik (BSI) bietet auf seiner Webseite [22b] detaillierte Informationen zu den verschiedenen Arten von Deepfakes an. Für den Anwendungsfall von Deepfakes im VideoIdent-Verfahren ist es theoretisch ausreichend, mittels Deepfakes ein Face Reenactment zu erreichen.

2.1 Zugrundeliegende Technologien

Die Erstellung von Deepfake-Videos ist ein Resultat der intensiven Forschung im Bereich der Computergrafik. Entgegen des ursprünglichen Vorgehens in der Computergrafik, die das konkrete Modellieren von Szenen umfasst, wird hier ein statistisches Lernen aus realem Bildmaterial durch Verfahren des maschinellen Lernens umgesetzt. Wie auch schon vom Begriff „Deepfake“ angedeutet, kommen hier Techniken aus dem Bereich des *deep learnings* zum Einsatz. Dabei handelt es sich um eine Unterklasse des maschinellen Lernens, die auf mehrschichtigen neuronalen Netzen basiert (im Detail bei LeCun et al. [LBH15]). Diese neuronalen Netze imitieren die Funktionsweise des Gehirns und bestehen dementsprechend aus einer Menge an untereinander verknüpften Neuronen, die in aufeinander folgenden Schichten oder auch *layers* angeordnet sind. Schichten, welche die Eingaben des Algorithmus' erhalten, werden *input layers* genannt, Schichten, welche das Ergebnis liefern, sind *output layers*. Dazwischenliegende Schichten, die nicht von außen ersichtlich sind, heißen *hidden layers*. Die Neuronen in den layers führen dabei die Berechnungen aus und geben ihre Ergebnisse an die mit ihnen verbundenen Neuronen weiter. Die Parameter der Berechnungen lernen sie dabei ohne das Zutun von Menschen (dazu ausführlich LeCun et al. [LBH15]). Im Bereich der Computergrafik haben sich neuronale Netze vor allem aufgrund der Fortschritte um *Convolutional Neural Networks (CNNs)* und *Generative Adversarial Networks (GANs)* durchgesetzt (näheres bei Tewari et al. [Te20]). CNNs ermöglichen es, mehrere zusammenhängende Datenpunkte (etwa

nebeneinanderliegende Pixel) gemeinsam zu betrachten und somit Kontextinformationen zu erfassen. So können Linien oder auch komplexe Objekte erkannt werden. Ursprünglich waren CNNs jedoch nur auf Klassifikationen ausgelegt und konnten keine neuen Inhalte generieren. Daher werden sie in der Computergrafik zumeist in der Kombination mit den von Vaswani et al. [Va17] vorgeschlagenen *Transformers* und GANs verwendet. GANs beschreiben ein Vorgehen für einen insgesamten Lernprozess. Sie wurden erstmals von Goodfellow et al. [Go14] vorgestellt und beschreiben eine Art und Weise, wie (content-) generierende neuronale Netze erstellt werden können. Das Vorgehen umfasst dabei einen Generator, der den neuen Content (wie bspw. ein Deepfake-Video) erstellt und einen Diskriminator, der versucht, zwischen echtem und generiertem Content zu unterscheiden. Durch stufenweises abwechselndes Verbessern der Parameter versucht der Diskriminator einerseits, immer besser zwischen echten und unechten Daten zu unterscheiden, und der Generator andererseits, den Diskriminator bestmöglich zu überlisten.

2.2 Deepfake-Videos im VideoIdent-Verfahren

Dass Deepfake-Videos von hoher Qualität auch außerhalb der Forschung weit verbreitet sind, haben Pu et al. [Pu21] in ihrer Studie belegt. Dort überprüfen sie auch die Performance von State-of-the-Art-Methoden zum Erkennen von Deepfakes auf von Laien erstellten Videos. Sie stellen fest, dass diese Methoden nur mäßig erfolgreich sind. Hinzu sollte hier bedacht werden, dass in dieser Studie nur Deepfake-Videos betrachtet wurden, die offensichtlich als solche gekennzeichnet waren. Damit ist nicht auszuschließen, dass eine Reihe von Videos von noch höherer Qualität unerkannt blieb, da diese Videos nicht ausgezeichnet waren und vielleicht auch nicht als Deepfake erkennbar waren. Ihre Studie beweist jedoch, dass theoretisch die Möglichkeit für Laien besteht, echt wirkende Deepfake-Videos ohne großen Aufwand zu erstellen und zu nutzen. Bekannte Beispiele sind das Open-Source-Projekt DeepFaceLab [22a], welches aufgrund seiner Bandbreite an Möglichkeiten recht komplex ist, und Zao [22d] als mobile Anwendung, die recht einfach zu benutzen ist. Wichtig ist anzumerken, dass diese Anwendungen noch keine Deepfakes in Echtzeit produzieren können. Nichtsdestotrotz zeigt ihre weite Verbreitung, dass derartige Software von Personen ohne detailliertes Wissen über die Erstellung von Deepfakes genutzt werden kann und die Möglichkeit des Erstellen von Deepfake-Videos in Echtzeit lediglich davon abhängt, wann eine derartige Anwendung der breiten Masse zur Verfügung gestellt wird. Als Lernmaterial und somit Voraussetzung für die Erstellung von Deepfake-Videos benötigen diese Verfahren teilweise nur ein kurzes Video, wie bei Kim et al. [Ki18] oder auch Thies et al. [Th18]. Andere Verfahren schaffen es sogar, nur anhand von ein paar oder auch nur eines einzigen Fotos realistische Videos zu erstellen wie bspw. bei Zakharov et al. [Za19] und Averbuch-Elor et al. [Av17] zu sehen. Zudem wird für diese Ergebnisse keine besonders teure oder seltene Hardware benötigt, wie Thies et al. ([Th16], [Th18]) zeigt. Dort wird auch bewiesen, dass Verfahren zur Erstellung von qualitativ hochwertigen Deepfakes mit handelsüblicher Hardware in Echtzeit möglich ist. Kim et al. [Ki18] zeigen zudem, dass nicht nur der Kopf, sondern bspw. auch der komplette Oberkörper manipuliert werden kann.

Deepfake-Videos von Personen und vor allem deren Köpfen haben jedoch auch Grenzen in ihrer Realitätsnähe. Während die hier vorgestellten Verfahren die ursprünglichen Probleme der korrekten Darstellung der Blickrichtung (bspw. bei Kim et al. [Ki18]) und des Mundes (z. Bsp. bei Thies et al. [Th16]) lösen, bleiben immer noch einige Aspekte der Ungenauigkeit in den erstellten Videos. Unter anderem seltene oder außergewöhnliche Körperhaltungen führen zu unrealistischen Ergebnissen und die Darstellung von Hautfalten bspw. an den Mundwinkeln erfolgt oftmals noch nicht vollkommen korrekt (bspw. bei [Th18] zu sehen). Andererseits können auch (Bart-) Haare zu einer Verschlechterung der Ergebnisse führen (exemplarisch bei [Th16] gezeigt). Zudem kann die erzeugte Auflösung der Videos auch von der Trainingszeit und der zur Verfügung stehenden Rechenleistung beschränkt werden, wie bei Kim et al. [Ki18] zu sehen. In Anbetracht des dauerhaften Fortschritts der Technik ist jedoch nicht auszuschließen, dass diese kleineren Probleme ebenfalls bald gelöst werden können. Bei der Nutzung von Deepfakes im VideoIdent-Verfahren kann aber auch eine Zielperson gewählt werden, die für diese Dinge weniger anfällig ist (bspw. kurze Haare oder kein Bart). Zudem könnte eine verminderte Videoqualität im Kontext eines Videoanrufs sowie weitere Kleinigkeiten (bspw. ungenaue Darstellung von Hautfalten) auch mit einer schlechten Internetverbindung begründet werden. Weiterhin ist wohl davon auszugehen, dass Videoanrufe im VideoIdent-Verfahren größtenteils ohne technische Unterstützung zur Erkennung von Deepfakes durchgeführt werden und es somit alleine den Mitarbeitern der VideoIdent-Anbieter obliegt, Deepfake-Videos zu erkennen. Betrachten diese die Videos nicht ausgiebig oder detailliert genug, auf einem zu kleinen Fenster oder wissen nicht, welche Anzeichen auf ein Deepfake-Video hinweisen, ist nicht zu erwarten, dass effektiv Fälschungen dieser Art erkannt werden. Die Tatsache, dass Deepfake-Videoanrufe auf dem heutigen Stand der Technik tatsächlich Personen täuschen können, hat u.a. das BSI [21] betont, indem es bestätigt, dass im Berichtszeitraum mehrere europäische Politiker einer Täuschung durch Deepfakes in einem Videoanruf zum Opfer gefallen sind, und auch die Agentur der Europäischen Union für Cybersicherheit (ENISA) [22c] sieht die Gefahr von Deepfakes in der Videoidentifikation als realistisch an. Anzumerken ist hier, dass für die Authentifizierung im VideoIdent-Verfahren ordnungsgemäß ebenfalls der Personalausweis vor der Kamera gezeigt werden muss. Der ENISA [22c] zufolge ist jedoch auch das digitale Fälschen des Personalausweises mitsamt der Lichtspiegelungen und Muster, die das Fälschen des physischen Personalausweis verhindern sollen, durch Deepfakes möglich.

3 Rechtliche Einordnung

Wie gesehen, handelt es sich bei der Nutzung von Deepfakes bei der Authentifizierung um ein realistisches Szenario. Fraglich sind jedoch die rechtlichen Konsequenzen.

3.1 Verletzung des allgemeinen Persönlichkeitsrechts

In Betracht kommt eine Verletzung des allgemeinen Persönlichkeitsrechts (APR). Dieses findet im Rahmen mittelbarer Drittwirkung bei Streitigkeiten zwischen Privatpersonen

über zivilrechtliche Generalklauseln Anwendung ([Gs22], *Hermann*, § 823 BGB Rn. 1659). Das APR steht nicht explizit im Grundgesetz der Bundesrepublik Deutschland, sondern wird aus Art. 2 Abs. 1 GG in Verbindung mit Art. 1 Abs. 1 GG hergeleitet ([He21], *Di Fabio*, Art. 2 GG Rn. 128). Gewährleistet und geschützt wird durch das APR die engere persönliche Lebenssphäre ([JK20], *Jarass*, Art. 2 GG, Rn. 36a). Der Schutzbereich des APRs als von der Rechtsprechung entwickeltes Recht ergibt sich aus der Rechtsprechung des Bundesverfassungsgerichts (BVerfG) ([MPS22], *Schmidt*, Art. 2 GG Rn. 34). Laut dem BVerfG ist der Schutzbereich nicht abschließend umschrieben, sondern es werden Fallgruppen entwickelt ([Bu86], 1860). Eine der Fallgruppen stellt das Recht am eigenen Bild dar; so hat jeder das Recht, selbst darüber zu entscheiden, wie er in der Öffentlichkeit dargestellt werden will ([Bu05], 3272). Der Einzelne erhält Einfluss- sowie Entscheidungsmöglichkeiten hinsichtlich der Erstellung und Verbreitung von Fotografien durch andere Personen ([Bu00], 1022), wobei auch der Schutz vor der Verbreitung von technisch manipulierten Bildern, die den Anschein eines authentischen Abbildes einer Person erwecken, umfasst ist ([Bu05], 3272). Die Herstellung von Deepfakes berührt daher den Schutzbereich des APRs. Dieser ist verletzt, soweit (wie im vorliegenden Fall) der Eingriff nicht gerechtfertigt werden kann. Das Recht am eigenen Bild ist auch einfachgesetzlich in den §§ 22, 23 KUG geregelt (vgl. hierzu: Abschnitt 3.2.4). Eine weitere Fallgruppe des APRs ist das Namensrecht. Da der Name einer Person Ausdruck ihrer Individualität und Identität ist, kann sie verlangen, dass er durch die Rechtsordnung geschützt wird ([Bu07], 671). Das Namensrecht besitzt ebenfalls eine zivilrechtliche Anspruchsgrundlage: § 12 BGB ([St21], *Mansel*, § 12 BGB Rn. 1,vgl. hierzu Abschnitt 3.2.1). Neben dem Recht am eigenen Bild sowie dem Namensrecht ist an dieser Stelle noch auf das Recht auf informationelle Selbstbestimmung zu verweisen, wonach auch das Recht darauf, selbst über die Preisgabe und Verwendung personenbezogener Daten zu entscheiden, vom APR geschützt wird ([Bu84], 422).

3.2 Zivilrechtliche Ansprüche

Im Folgenden werden die zivilrechtlichen Ansprüche geprüft, die die Person, deren Aussehen imitiert wurde, gegenüber dem Täter geltend machen kann.

3.2.1 Anspruch aus der Verletzung des Namensrechts, § 12 BGB

Die erste Rechtsgrundlage, aus der sich ein Unterlassungs- oder Beseitigungsanspruch ergeben könnte, ist § 12 BGB (Namensrecht). Verletzt ist das Namensrecht gemäß § 12 BGB dann, wenn die Rechte des Berechtigten durch unbefugten Gebrauch seines Namens durch einen anderen verletzt werden, § 12 S. 1 BGB (unbefugte Namensanmaßung). Da sich der Verwender des Deepfakes hier der Identität einer anderen Person ohne Nutzungsrecht bedient, ist der Anspruch gemäß § 12 BGB gegeben.

3.2.2 Unterlassungsanspruch aus § 1004 BGB in Verbindung mit dem APR

Auch § 1004 BGB gewährt als quasi-negatorischer Abwehranspruch Beseitigungs- und Unterlassungsansprüche für die Beeinträchtigung absoluter Rechte ([SS19], *Volkmann*, § 1004 BGB Rn. 1). Zu den *absoluten Rechten* in diesem Sinne gehört insbesondere auch das oben beschriebene APR und damit auch das Recht am eigenen Bild ([Bu06], 208). Die Verletzung des Rechts am eigenen Bild führt bei Vorliegen der sonstigen Voraussetzungen des § 1004 BGB zu einem Unterlassungs- bzw. Beseitigungsanspruch ([Ga20], *Raff*, § 1004 BGB Rn. 37 f.).

3.2.3 Schadensersatzanspruch aus § 823 Abs. 1 BGB

§ 823 Abs. 1 BGB ist ein deliktischer Anspruch und gewährt Schadensersatz für Rechtsgutsverletzungen, die durch eine unerlaubte Handlung eingetreten sind. Aus dem Wortlaut des § 823 Abs. 1 GG geht hervor, dass für widerrechtliche Verletzungen der in Abs. 1 explizit aufgezählten Rechtsgüter oder von sonstigen Rechten Schadensersatz zu zahlen ist. Zu den *sonstigen Rechten* gehören u.a. einzelne Persönlichkeitsrechte ([St21], *Teichmann*, § 823 BGB, Rn. 12), wozu das Recht am eigenen Bild (§ 22 KUG) sowie das Namensrecht (§ 12 BGB) zählen ([St21], *Teichmann*, § 823 BGB, Rn. 13; [Ha20], *Wagner*, § 823 BGB, Rn. 416; [HP22], *Förster*, § 823 BGB, Rn. 159). Um einen Anspruch aus § 823 Abs. 1 BGB zu begründen, bedarf es einer rechtswidrigen Verletzung eines dieser Rechte durch eine vorsätzliche oder fahrlässige Verletzungshandlung ([Gs22], *Spindler*, § 823 BGB, Rn. 70 ff.). Im Falle der Nutzung von Deepfakes bei der Authentifizierung im VideoIdent-Verfahren liegt eine vorsätzliche Verletzung des Rechts am eigenen Bild sowie des Namensrechts vor. Einzig fraglich ist hier die *Rechtswidrigkeit* der Verletzungshandlung. Anders als bei der Verletzung der in § 823 Abs. 1 BGB ausdrücklich genannten Rechte, wird bei der Verletzung des APRs die Rechtswidrigkeit nicht durch die Tatbestandsmäßigkeit indiziert ([Gs22], *Specht-Riemenschneider*, § 823 BGB, Rn. 1461). Der Eingriff in den Schutzbereich des APRs ist nur dann rechtswidrig, wenn die schutzwürdigen Interessen des Verletzten überwiegen ([Bu14], 696 Rn. 22) (positive Feststellung der Rechtswidrigkeit, ([Gs22], *Specht-Riemenschneider*, § 823 BGB, Rn. 1461 f.)). In dem hier betrachteten Fall fällt diese Abwägung eindeutig zugunsten des Verletzten aus: Dessen Recht am eigenen Bild bzw. Namensrecht wurde aus verwerflichen Motiven verletzt. Die Interessen des Deepfakes-Verwenders sind indes schon gar nicht schutzwürdig, weshalb dem Verletzten hier ein Anspruch auf Schadensersatz zusteht.

3.2.4 Schadensersatzanspruch aus einer Schutzgesetzverletzung, § 823 Abs. 2 BGB

Ein Schadensersatzanspruch könnte sich auch aus § 823 Abs. 2 BGB ergeben. § 823 Abs. 2 BGB fordert die zurechenbare Verletzung eines sogenannten Schutzgesetzes

([Gs22], *Spindler*, § 823 BGB, Rn. 255). Ein Schutzgesetz ist gemäß § 823 Abs. 2 S. 1 BGB ein Gesetz, welches den Schutz eines anderen bezweckt. Als Schutzgesetz kommen hier sowohl § 22 KUG, als auch möglicherweise verwirklichte Straftatbestände in Frage. Da die Verwirklichung der Straftatbestände unter Abschnitt 3.3 geprüft wird, soll der Fokus dieses Abschnitts auf § 22 KUG liegen. Das KUG wird vorliegend aufgrund der Öffnungsklausel aus Art. 85 Abs. 1 DSGVO als anwendbar betrachtet, obwohl das Verhältnis von DSGVO und KUG sich weiterhin im Diskurs befindet (Vgl. hierzu:[KW19]). § 22 KUG ist als Schutzgesetz in diesem Sinne anerkannt ([HP22], *Förster*, § 823 BGB, Rn. 291 m.w.N.). Laut § 22 KUG dürfen Bildnisse nur mit Einwilligung des Abgebildeten verbreitet oder öffentlich zur Schau gestellt werden. Bildnisse sind Darstellungen von Personen, die deren äußere Erscheinung in einer für Dritte erkennbaren Weise wiedergeben ([LLO20], *Götting*, § 22 KUG Rn. 14). Im vorliegenden Fall bestehen keine Zweifel daran, dass ein Bildnis verwendet wird: Der Täter will den Eindruck erwecken, dass die abgebildete Person handelt. Deren Darstellung ist also das ausgemachte Ziel des Täters. Problematisch erscheint jedoch die erforderliche Handlung, um § 22 KUG zu erfüllen. § 22 KUG verbietet nämlich nur die Verbreitung sowie die öffentliche Zurschaustellung des Bildnisses; die schlichte Herstellung fällt nicht unter § 22 KUG. Zwar wird vom „Verbreiten“ jegliche Art der Verbreitung erfasst. Jedoch ist erforderlich, dass der Verbreitende die Kontrolle darüber verliert, inwieweit das Bildnis an die Öffentlichkeit gelangt ([LLO20], *Götting*, § 22 KUG, Rn. 36). Die öffentliche Zurschaustellung ist die Sichtbarmachung des Bildnisses im weitesten Sinne ([LLO20], *Götting*, § 22 KUG Rn. 37), das Bildnis muss der Öffentlichkeit zur Verfügung gestellt werden. Beide Varianten erscheinen hier nicht erfüllt. Wird das Foto lediglich gegenüber dem Angestellten einer Bank zur Authentifizierung verwendet, handelt es sich zwar durchaus um eine Verletzung des APRs — § 22 KUG ist hingegen nicht verletzt. Damit einher geht auch, dass eine Strafbarkeit nach § 33 KUG nicht in Betracht kommt.

3.3 Strafrechtliche Aspekte

Bei der strafrechtliche Bewertung ist zwischen Straftaten, die an die Nutzung des Deepfakes im VideoIdent-Verfahren als Tathandlung anknüpfen, und denen, die anschließendes Verhalten unter Strafe stellen, zu unterscheiden.

3.3.1 Betrug, § 263 StGB

In Betracht käme vorliegend die Verwirklichung eines Betrugs gemäß § 263 Abs. 1 StGB. Der Betrug könnte hier sowohl *gegenüber* einem Bankangestellten und *zulasten* einer dritten Person (deren Identität verwendet wurde) als auch *gegenüber* eines Bankangestellten und *zulasten* der Bank verwirklicht worden sein. Ein Betrug liegt vor, wenn durch die Vorspiegelung falscher Tatsachen oder die Entstellung oder Unterdrückung wahrer Tatsachen bei einer anderen Person ein Irrtum erregt oder unterhalten wird und dadurch das Vermögen eines anderen beschädigt wird. Dabei muss der Täter in der Absicht handeln, sich oder

einem Dritten einen rechtswidrigen Vermögensvorteil zu verschaffen. Mit einer Täuschung über Tatsachen möchte der Täter auf die Vorstellungskraft des Gegenübers einwirken und ihn dadurch in die Irre führen, wobei die Täuschung sowohl durch positives Tun (auch durch schlüssiges Verhalten) als auch durch Unterlassen verwirklicht werden kann ([LHK18], *Kühl*, § 263 StGB Rn. 6 ff.). Zu *Tatsachen* im Sinne des § 263 StGB gehören alle Zustände und Geschehnisse der Vergangenheit und der Gegenwart, die dem Beweis zugänglich sind ([KNP17], *Kindhäuser*, § 263 StGB Rn. 73), auch die Identität einer Person ([He22a], *Hefendehl*, § 263 StGB, Rn. 99). Setzt eine Person also im VideoIdent-Verfahren Deepfakes ein, um über ihre Identität zu täuschen, ist jedenfalls eine tatbestandliche Täuschung gegeben. Soweit das Gegenüber von der Richtigkeit der vorgetäuschten Identität ausgeht, erliegt es auch dem erforderlichen Irrtum, also einer Fehlvorstellung über Tatsachen ([Sc19], *Perron*, § 263 StGB Rn. 32a). Problematisch könnte vorliegend jedoch das ungeschriebene Tatbestandsmerkmal der Vermögensverfügung sein. Eine Vermögensverfügung wird als jedes Tun, Dulden oder Unterlassen des Getäuschten definiert, welches sich *unmittelbar* vermögensmindernd beim Getäuschten oder einem Dritten auswirkt ([Bu60], 1069). Dabei ist zwingend erforderlich, dass Getäuschter und Verfügender identisch sind. („Dreiecksbetrug“ ([He22b], *Beukelmann*, § 263 Rn. 34)). Hier ist jedoch fraglich, ob überhaupt eine Vermögensverfügung vorliegt. Selbst wenn nach erfolgter Authentifizierung der Zugriff auf ein Bankkonto ermöglicht würde, wird hiermit noch keine Vermögensminderung ausgelöst. Zu einer Minderung des Vermögens käme es erst durch ein weiteres Verhalten des Täters — wenn dieser bspw. den Dispokredit überzieht — *und* der Bank - wenn diese sich an den vermeintlichen Kontoinhaber wegen etwaiger Ansprüche wendet. Der Betrug ist ein *Selbstschädigungsdelikt*, was bedeutet, dass die Vermögensschädigung unmittelbar durch ein Handeln des Opfers ([He22b], *Beukelmann*, § 263 StGB, Rn. 32) erfolgt. Die Unmittelbarkeit ist zu verneinen, wenn durch die Handlung des Getäuschten lediglich Zugriffsmöglichkeiten des Täters erhöht werden und dieser damit die Möglichkeit erhält, durch eigenständige Schritte eine Vermögensverschiebung auszulösen ([He22b], *Beukelmann*, § 263 StGB Rn. 32). Zwar können auch mehraktige Geschehen zu einem Betrug führen, essentiell ist jedoch, dass die Handlung, durch die die Vermögensverschiebung ausgelöst wird, auch vom Getäuschten vollzogen wird ([KNP17], *Kindhäuser*, § 263 StGB, Rn. 202). Dies ist hier jedoch nicht der Fall. Erforderlich sind weitere Handlungen des Täters selbst, was schon dem Selbstschädigungscharakter des Betruges widerspricht. Ein Betrug und damit auch ein Schadensersatzanspruch aus § 823 Abs. 2 BGB i.V.m. § 263 StGB sind mithin nicht verwirklicht.

3.3.2 Geldwäsche, § 261 StGB

Ferner könnte hier der Straftatbestand der Geldwäsche verwirklicht worden sein. Gemäß § 261 Abs. 1 StGB macht sich strafbar, wer einen Gegenstand, der aus einer in § 261 Abs. 1 S. 2 StGB näher bezeichneten rechtswidrigen Tat herrührt, verbirgt, dessen Herkunft verschleiert oder die Ermittlung der Herkunft, das Auffinden, die Einziehung oder die Sicherstellung eines solchen Gegenstandes vereitelt oder gefährdet. Es bedarf mithin

eines *Gegenstandes*, der aus einer rechtswidrigen Tat herrührt. Unter einem Gegenstand in diesem Sinne ist jeder Vermögensgegenstand, der seinem Inhalt nach bewegliche oder unbewegliche Sachen oder Rechte umfasst, zu verstehen ([Bu15], 3254 Rn. 4). Ausgangspunkt des Gegenstandsbegriff ist § 90 BGB, wonach Sachen nur körperliche Gegenstände sind ([Sa21], *Neuheuser*,§ 261 StGB Rn. 35). Durch den Deepfake-Einsatz im VideoIdent-Verfahren kann zwar ein Konto eröffnet werden, einen vermögenswerten Gegenstand erhält der Täter dadurch jedoch wohl nicht. Selbst wenn man den Erhalt eines vermögenswerten Gegenstandes aufgrund des Erhalts von Zugangsdaten zu einem Konto bejahen wollte, so mangelt es an der Katalogstraftat gemäß § 261 Abs. 1 S. 2 StGB, aus der dieser „Gegenstand" herrührt. Der Täter macht sich hier also nicht der Geldwäsche gemäß § 261 Abs. 1 StGB schuldig.

3.3.3 Verletzung des höchstpersönlichen Lebensbereichs und von Persönlichkeitsrechten durch Bildaufnahmen, § 201a StGB

Im Rahmen der rechtlichen Bewertung der Verwendung von Deepfakes wird häufig auf die Strafnorm des § 201a StGB verwiesen ([La19], 576; [Th21], 204; [La20], 79), welche die Verletzung des höchstpersönlichen Lebensbereichs und von Persönlichkeitsrechten durch Bildaufnahmen unter Strafe stellt. Die Herstellung des Deepfakes ist von § 201a StGB indes nicht erfasst: Die Nachbearbeitung von Bildern bzw. computergenerierte Bilder fallen nicht unter den Begriff der „Bildaufnahme" in diesem Sinne ([Sc19], *Eisele*, § 201a StGB, Rn. 6; [La20], 79), womit eine Strafbarkeit nach § 201a Abs. 1 StGB ausscheidet. Jedoch ist § 201a Abs. 2 StGB regelmäßig bei der Verbreitung pornographischer Deepfakes erfüllt ([La20], 79). Dieser Fall unterscheidet sich jedoch vom hier betrachteten Szenario: Hier werden keine Fotos verbreitet, die dazu geeignet sind, das Ansehen des Abgebildeten erheblich zu beeinträchtigen. Dies ist bei einem pornographischen Deepfakes naturgemäß anders zu bewerten. Werden Deepfakes jedoch zur Authentifizierung im VideoIdent-Verfahren genutzt, kommt keine Strafbarkeit nach § 201a Abs. 2 StGB in Betracht.

3.3.4 Datenschutzrechtliche Strafvorschrift, § 42 Abs. 2 BDSG

Neben den angeprüften und abgelehnten Delikten kommt für die Handlung des Täters ferner die Verletzung datenschutzrechtlicher Vorschriften und damit einhergehend die Erfüllung der datenschutzspezifischen Strafnorm des § 42 BDSG in Betracht. Zu beachten ist, dass es sich gemäß § 42 Abs. 3 S. 1 BDSG um ein absolutes Antragsdelikt handelt ([KB20], *Bergt*, § 42 BDSG Rn. 56). Eine Tat im Sinne des § 42 Abs. 1 oder Abs. 2 BDSG wird also stets nur auf Antrag der betroffenen Person, des Verantwortlichen, des Bundesbeauftragten oder der Aufsichtsbehörde verfolgt. Erfüllt sein könnte hier der § 42 Abs. 2 Nr. 1 BDSG. § 42 Abs. 1 BDSG ist indes nicht gegeben, da die verwendeten Daten weder einer großen Anzahl an Personen zugänglich gemacht werden, noch ohne Weiteres von einem gewerbsmäßigen Handeln des Täters auszugehen ist. § 42 Abs. 2 Nr. 1 BDSG stellt die unrechtmäßige

Verarbeitung nicht allgemein zugänglicher personenbezogener Daten gegen ein Entgelt oder in der Absicht, sich oder einen anderen zu bereichern oder einen anderen zu schädigen, unter Strafe. Personenbezogene Daten sind, im Einklang mit Art. 4 Nr 1 DSGVO ([BW22], *Brodowski/Nowak*, § 42 BDSG Rn. 22), alle Informationen, die sich auf identifizierte oder identifizierbare Personen beziehen. Hierzu zählen bspw. auch die Gesichtszüge einer Person ([La20], 80). Der Begriff der Datenverarbeitung nach Art. 4 Nr. 2 DSGVO geht indes sehr weit: So ist jegliches datenschutzrelevantes Verhalten hiervon erfasst ([KB20], *Bergt*, § 42 BDSG Rn. 32). Daher fällt hierunter sowohl die Speicherung des Ausgangsmaterials als auch die Deepfake-Herstellung und auch -Verwendung ([La20], 80). Da im vorliegenden Fall auch von der erforderlichen Bereicherungsabsicht zu eigenen Gunsten auszugehen ist, erscheint die Verwirklichung von § 42 Abs. 2 Nr. 1 BDSG hier durchaus denkbar. Insbesondere ist nicht erforderlich, dass sich der Vermögensvorteil unmittelbar aus der Tathandlung ergibt ([KB20], *Bergt*, § 42 BDSG, Rn. 49). Einzig problematisch erscheint das Tatbestandsmerkmal der Zugänglichkeit. Daten sind im Sinne des § 42 Abs. 2 Nr. 1 BDSG öffentlich zugänglich, wenn sie von jedermann wahrgenommen werden können, ohne dass es rechtliche Beschränkungen gibt ([KB20], *Bergt*, § 42 BDSG Rn. 8). Nicht öffentlich zugänglich sind beispielsweise Fotos, die auf dem privaten Social-Media-Profil geteilt werden ([La20], 80). Anders ist dies jedoch zu bewerten, wenn das Social-Media-Profil ohne Anmeldung öffentlich abrufbar ist ([KB20], *Bergt*, § 42 BDSG Rn. 9). Nimmt der Täter daher Fotos und den Namen, die auf einer technisch nicht besonders geschützten Webseite ([KB20], *Bergt*, § 42 BDSG Rn. 9) veröffentlicht sind, entzieht er sich der Strafbarkeit. Hinweise auf der Webseite, dass die Fotos nicht verwendet werden dürfen oder dass der Zugriff bestimmten Personen vorbehalten ist, sind dabei unbeachtlich ([KB20], *Bergt*, § 42 BDSG Rn. 9; [BW22], *Brodowski/Nowak*, § 42 BDSG Rn. 26). Betrachtet man die Fülle an teils hochwertigen Fotos auf öffentlich und ohne Anmeldung verfügbaren *Instagram*-Profilen, dürfte es ein leichtes Unterfangen sein, öffentlich zugängliche Fotos zu verwenden. Werden hingegen private Fotos und Informationen genutzt, ist eine Strafbarkeit gemäß § 42 Abs. 2 Nr. 1 BDSG gegeben, wodurch auch ein Anspruch auf Schadensersatz in Verbindung mit § 823 Abs. 2 BGB entsteht.

3.3.5 Urkundenfälschung, § 267 StGB, Datenveränderung, § 303a StGB und Vorbereiten des Ausspähens und Abfangens von Daten, § 202c StGB

Da ein Personalausweis als Urkunde anzusehen ist ([Sc19], *Heine/Schuster*, § 267 StGB Rn. 44), könnte eine Urkundenfälschung, § 267 StGB, in der Veränderung von Name und Lichtbild auf dem im VideoIdent-Verfahren vorgezeigten Ausweis liegen. Es ist jedoch zu beachten, dass durch die vorübergehende Bearbeitung des Personalausweises kein physisches Dokument hergestellt wird, somit zu keinem Zeitpunkt eine unechte Urkunde vorliegt und mithin auch keine Urkundenfälschung (siehe hierzu: [Bu11], 214). Einer Datenveränderung nach § 303a StGB macht sich schuldig, wer rechtswidrig Daten (§ 202a Abs. 2) löscht, unterdrückt, unbrauchbar macht oder verändert. Unter der *Veränderung* wird in diesem Sinne jede inhaltliche Umgestaltung von gespeicherten Daten, die mit einer

Funktionsbeeinträchtigung verbunden ist ([HH19], *Wieck-Noodt*, § 303a StGB Rn. 15), verstanden. Damit erscheint das Handeln des Täters im hier beschriebenen Szenario bereits nicht tatbestandsmäßig: So werden hier keine *gespeicherten* Daten verändert. Auch bei Nutzung eines bereits existierenden Bildes des Opfers wird der Informationsgehalt der Daten nicht abgeändert. Eine Datenveränderung nach § 303a Abs. 1 StGB liegt daher nicht vor. § 202a StGB stellt das Ausspähen von Daten, § 202c StGB entsprechende Vorbereitungshandlungen unter Strafe. Da der Täter im Erfolgsfall ein Passwort erhalten könnte, welches auf den Namen einer anderen Person läuft, könnte § 202c Abs. 1 Nr 1 StGB verwirklicht sein. Jedoch bleibt zu beachten, dass der Täter selbst wenn er ein Passwort oder Online-Banking-Zugangsdaten erhalten sollte, keinen Zugang zu Daten einer anderen Person erhält, womit auch diese Straftatbestände ausscheiden.

3.3.6 Strafbarkeit anschließender Handlungen des Täters

Betrachtet man als Tathandlung nicht die Verwendung der Deepfakes im VideoIdent-Verfahren, sondern anschließende Handlungen des Täters, so kommt eine Strafbarkeit wegen Betruges, § 263 StGB oder wegen Computerbetruges, § 263a StGB in Betracht. Die 3. Variante des § 263a StGB (die unbefugte Verwendung von Daten) könnte der Verwender dann verwirklichen, wenn er das eröffnete Konto bspw. durch Einkäufe in Onlineshops unter Nutzung von Online-Banking-Daten überzieht. Um einen Computerbetrug zu realisieren, müsste er durch die unbefugte Verwendung von Daten einen Datenverarbeitungsvorgang beeinflussen und dadurch das Vermögen einer anderen Person schädigen ([He22a], *Hefendehl/Noll*, § 263a StGB Rn. 20). Die unbefugte Verwendung von Daten liegt nach der herrschenden betrugsnahen Auslegung des § 263a Abs. 1 Var. 3 StGB vor, wenn der Berechtigte oder ein Dritter getäuscht würden, wenn der Täter jene Daten ihm gegenüber verwenden würde ([MR20], *Altenhain*, § 263a StGB Rn. 12). Da der Täter sich als andere Person ausgibt und damit über seine Identität täuscht, kann man von einer Täuschung ausgehen. Die Überziehung eines fremden Kontos durch Online-Bestellungen ist auch einer der anerkannten Fälle des Computerbetrugs ([MR20], *Altenhain*, § 263a StGB Rn. 16). Geschädigt ist in diesen Fällen dann regelmäßig der Berechtigte. Einen Betrug könnte der Täter realisieren, wenn er eine erhaltene Bankkarte verwendet, um Geld abzuheben oder um eine Zahlung zu tätigen. Im vorliegenden Fall besteht jedoch ein Unterschied zu sonst gängigen Fällen: Zwischen der Bank und dem „Berechtigten“, also der Person, deren Identität verwendet wurde, gab es nie Kontakt. Mangels wirksamer Stellvertretung durch den Täter kann man hier auch nicht davon ausgehen, dass eine Vertragsbeziehung zustande gekommen ist. Da für eine Bank der Name und die Identität des Gegenübers von Relevanz ist, ist auch kein Vertrag mit dem Täter selbst zustande gekommen. Ein Schaden des Berechtigten liegt mithin gar nicht vor: Da sein eigenes Konto nicht überzogen wurde und auch keine vertragliche Beziehung zwischen ihm und der Bank zustande kam, treffen ihn keine Ansprüche und sein Vermögen ist nicht gemindert. In Frage kommt also nur ein Betrug zulasten der Bank, die hier wohl mangels Ansprechpartner den erlittenen Schaden nicht ausgleichen kann. Ein Computerbetrug gemäß § 263a Abs. 1 Var. 3 StGB

könnte hier also ebenso wie ein Betrug nach § 263 StGB erfüllt sein und mit ihm auch ein Schadensersatzanspruch gemäß § 823 Abs. 2 BGB i.V.m. dem jeweiligen Straftatbestand.

4 Fazit

Im Verlaufe dieses Beitrags wurde aufgezeigt, dass Deepfakes Ergebnisse des deep learning sind und auf neuronalen Netzen basieren und ihre Produktion auch unter Laien weit verbreitet und leicht zugänglich ist. Zudem ist es technisch möglich, qualitativ hochwertige Deepfake-Videos in Echtzeit mithilfe weniger Fotos oder eines kurzen Videos zu erstellen, wobei keine besonders teure oder aufwändige Hardware benötigt wird. In Bezug auf die Anwendbarkeit im VideoIdent-Verfahren bedeutet dies, dass eine Nutzung von Deepfake-Videos zu unlauteren Zwecken ein durchaus realistisches Szenario darstellt. Rechtlich gesehen würde der Einsatz eines Deepfake-Videos im VideoIdent-Verfahren eine Verletzung des Allgemeinen Persönlichkeitsrechts der imitierten Person bedeuten, woraus Beseitigungs- und Unterlassungsansprüche sowie Schadensersatzansprüche resultieren. Es hat sich jedoch auch gezeigt, dass die strafrechtliche Einordnung des beschriebenen Szenarios deutlich komplexer ist: So scheitern mehrere Straftatbestände an einzelnen Tatbestandsmerkmalen, insbesondere ist mangels Vermögensverfügung kein Betrug gemäß § 263 StGB gegeben. Einzig der § 42 BDSG erscheint zumindest in Einzelfällen anwendbar, kann aber auch leicht umgangen werden. Es zeigt sich, dass der beim Deepfakes-Einsatz im VideoIdent-Verfahren vorgenommene „Identitätsdiebstahl“ als solcher keine strafrechtlichen Konsequenzen nach sich zieht, was als durchaus problematisch anzusehen ist. Bereits die Eröffnung eines Girokontos kann beispielsweise Auswirkungen auf die persönliche Schufa-Auskunft haben. Daran ändert auch die Tatsache, dass durch spätere Handlungen des Täters ein Computerbetrug verwirklicht werden könnte, nichts.

Dazu bleibt die Problematik des reinen Erkennens von Deepfake-Videos. Die technischen Methoden zur Erkennung von Deepfakes sind noch nicht sehr zuverlässig oder auf die breite Masse anwendbar, da sie stark von dem Erstellungsalgorithmus abhängen (im Detail bei Pu et al. [Pu21]). Auch eine breit angelegte *Deepfake Detection Challenge* [Do20] hat nur einen mäßigen Erfolg der teilnehmenden Algorithmen gezeigt. Dementsprechend bleibt die zuverlässige Erkennung von Deepfakes trotz ihrer weiten Verbreitung ein offenes Problem, welches durch den fehlenden rechtlichen Rahmen bei der Nutzung im VideoIdent-Verfahren nur noch unterstrichen wird und gemeinsam einen akuten Handlungsbedarf auf rechtlicher wie auch auf technischer Seite offenbart.

Literatur

[21] Die Lage der IT-Sicherheit in Deutschland 2021. Bundesamt für Sicherheit in der Informationstechnik (BSI), Bonn, 2021.

[22a] DeepFaceLab, 2022, URL: https://github.com/iperov/DeepFaceLab.

[22b] Deepfakes - Gefahren und Gegenmaßnahmen, 2022, URL: https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html.

[22c] Remote Identity Proofing - Attacks and Countermeasures, 2022, URL: https://www.enisa.europa.eu/publications/remote-identity-proofing-attacks-countermeasures.

[22d] ZAO, 2022, URL: https://www.zaoapp.net/.

[Av17] Averbuch-Elor, H.; Cohen-Or, D.; Kopf, J.; Cohen, M. F.: Bringing Portraits to Life. ACM Trans. Graph. 36/6, Nov. 2017, ISSN: 0730-0301, URL: https://doi.org/10.1145/3130800.3130818.

[Bu00] Bundesverfassungsgericht. Neue Juristische Wochenschrift/14, S. 1021–1026, 2000.

[Bu05] Bundesverfassungsgericht. Neue Juristische Wochenschrift/45, S. 3271–3273, 2005.

[Bu06] Bundesverfassungsgericht. Neue Juristische Wochenschrift/4, S. 207–211, 2006.

[Bu07] Bundesverfassungsgericht. Neue Juristische Wochenschrift/10, S. 671–672, 2007.

[Bu11] Bundesgerichtshof. Neue Zeitschrift für Strafrecht/7, S. 213–214, 2011.

[Bu14] Bundesgerichtshof. Gewerblicher Rechtsschutz und Urheberrecht/7, S. 693–702, 2014.

[Bu15] Bundesgerichtshof. Neue Juristische Wochenschrift/44, S. 3254–3255, 2015.

[Bu60] Bundesgerichtshof. Neue Juristische Wochenschrift/23, S. 1068–1069, 1960.

[Bu84] Bundesverfassungsgericht. Neue Juristische Wochenschrift/8, S. 419–428, 1984.

[Bu86] Bundesverfassungsgericht. Neue Juristische Wochenschrift/30, S. 1859–1861, 1986.

[BW22] Brink, S.; Wolff, H. A.: BeckOK Datenschutzrecht. C.H.Beck Verlag, München, 2022.

[Co] Commerzbank: Online-Legitimation, URL: https://www.commerzbank.de/portal/de/privatkunden/hilfe-kontakt/services/online-legitimation/online-legitimation.html, Stand: 19. 04. 2022.

[Do20] Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Ferrer, C. C.: The DeepFake Detection Challenge (DFDC) Dataset, 2020, URL: https://arxiv.org/abs/2006.07397.

[Ga20] Gaier, R.: Münchener Kommentar zum Bürgerlichen Gesetzbuch, Band 8, Sachenrecht §§ 854 - 1296, WEG - ErbbauRG. C.H.Beck Verlag, München, 2020.

[Go14] Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y.: Generative Adversarial Nets. In: Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2. NIPS'14, MIT Press, Montreal, Canada, S. 2672–2680, 2014.

[Gs22] Gsell, B.; Krüger, W.; Lorenz, S.; Reymann Christoph (Gesamthrsg. Zivilrecht); Artz, M.; Ball, W.; Benecke, M.; Geibel, S.; Gsell, B.; Hager, J.; Harke, J. D.; Kessen, M.; Köndgen, J.; Krafka, A.; Krüger, W.; Lobinger, T.; Looschelders, D.; Lorenz, S.; Maurer, H.-U.; Müller-Engels, G.; Reymann, C.; Schindler, W.; Schmidt, H.; Segna, U.; Sprickhoff, A.; Wellenhofer, M.: beck-online.Grosskommentar BGB. C.H.Beck Verlag, München, 2022.

[Ha20] Habersack, M.: Münchener Kommentar zum Bürgerlichen Gesetzbuch, Band 7, Schuldrecht - Besonderer Teil IV, §§ 705 - 853. C.H.Beck Verlag, München, 2020.

[He21] Herzog, R.; Scholz, R.; Herdegen, M.; Klein, H.: Grundgesetz Kommentar, Band I, Texte und Art. 1-5. C.H.Beck Verlag, München, 2021.

[He22a] Hefendehl, R.: Münchener Kommentar zum Strafgesetzbuch, Band 5 §§ 263 - 297 StGB. C.H.Beck Verlag, München, 2022.

[He22b] von Heintschel-Heinegg, B.: BeckOK StGB. C.H.Beck Verlag, München, 2022.

[HH19] Hefendehl, R.; Hohmann, O.: Münchener Kommentar zum Strafgesetzbuch, Band 5 §§ 263 - 358 StGB. C.H.Beck Verlag, München, 2019.

[HP22] Hau, W.; Poseck, R.: BeckOK BGB. C.H.Beck Verlag, München, 2022.

[IN] ING: Girokonto, URL: https://www.ing.de/girokonto/konto-eroeffnen/, Stand: 19. 04. 2022.

[JK20] Jarass, H.; Kment, M.: Grundgesetz für die Bundesrepublik Deutschland Kommentar. C.H.Beck Verlag, München, 2020.

[KB20] Kühling, J.; Buchner, B.: Datenschutz-Grundverordnung BDSG - Kommentar. C.H.Beck Verlag, München, 2020.

[Ki18] Kim, H.; Garrido, P.; Tewari, A.; Xu, W.; Thies, J.; Niessner, M.; Pérez, P.; Richardt, C.; Zollhöfer, M.; Theobalt, C.: Deep Video Portraits. ACM Trans. Graph. 37/4, Juli 2018, ISSN: 0730-0301, URL: https://doi.org/10.1145/3197517.3201283.

[KNP17] Kindhäuser, U.; Neumann, U.; Paeffgen, H.-U.: NomosKommentar Strafgesetzbuch. Nomos Verlag, Baden-Baden, 2017.

[KW19] Krüger, S.; Wiencke, J.: Bitte recht freundlich – Verhältnis zwischen KUG und DS-GVO Herstellung und Veröffentlichung von Personenbildnissen nach Inkrafttreten der DS-GVO. Multimedia und Recht/2, S. 76–80, 2019.

[La19] Lantwin, T.: Deep Fakes - Düstere Zeiten für den Persönlichkeitsschutz? Multimedia und Recht/9, S. 574–578, 2019.

[La20] Lantwin, T.: Strafrechtliche Bekämpfung missbräuchlicher Deepfakes - Geltendes Recht und möglicher Regelungsbedarf. Multimedia und Recht/2, S. 78–82, 2020.

[LBH15] LeCun, Y.; Bengio, Y.; Hinton, G.: Deep learning. Nature 521/, S. 436–444, 2015, URL: https://doi.org/10.1038/nature14539.

[LHK18] Lackner, K.; Heger, M.; Kühl, K.: Strafgesetzbuch Kommentar. C.H.Beck Verlag, München, 2018.

[LLO20] Loewenheim, U.; Leistner, M.; Ohly, A.: Urheberrecht Kommentar UrhG - KUG - VGG. C.H.Beck Verlag, München, 2020.

[MPS22] Müller-Glöge, R.; Preis, U.; Schmidt, I.: Erfurter Kommentar zum Arbeitsrecht, Band 51. C.H.Beck Verlag, München, 2022.

[MR20] Matt, H.; Renzikowski, J.: Strafgesetzbuch Kommentar. Verlag Franz Vahlen München, München, 2020.

[Pu21] Pu, J.; Mangaokar, N.; Kelly, L.; Bhattacharya, P.; Sundaram, K.; Javed, M.; Wang, B.; Viswanath, B.: Deepfake Videos in the Wild: Analysis and Detection. In: Proceedings of the Web Conference 2021. WWW '21, Association for Computing Machinery, Ljubljana, Slovenia, S. 981–992, 2021, ISBN: 9781450383127, URL: https://doi.org/10.1145/3442381.3449978.

[Sa17] Samantha Cole, Vice.com: AI-Assisted Fake Porn Is Here and We're All Fucked, 2017, URL: https://www.vice.com/en_us/article/gydydm/gal-gadot-fake-ai-porn, Stand: 29. 03. 2022.

[Sa21] Sander, G. M.: Münchener Kommentar zum Strafgesetzbuch, Band 4 §§ 185 - 262 StGB. C.H.Beck Verlag, München, 2021.

[Sc19] Schönke, A.; Schröder, H.; Eser, A.; Perron, W.; Sternberg-Lieben, D.; Eisele, J.; Hecker, B.; Kinzig, J.; Bosch, N.; Schuster, F.; Weißer, B.; Schittenhelm, U.: Schönke/Schröder Strafgesetzbuch Kommentar. C.H.Beck Verlag, München, 2019.

[SS19] Spindler, G.; Schuster, F.: Recht der elektronischen Medien Kommentar. C.H.Beck Verlag, München, 2019.

[St21] Stürner, R.: Jauernig Bürgerliches Gesetzbuch. C.H.Beck Verlag, München, 2021.

[Te20] Tewari, A.; Fried, O.; Thies, J.; Sitzmann, V.; Lombardi, S.; Sunkavalli, K.; Martin-Brualla, R.; Simon, T.; Saragih, J. M.; Nießner, M.; Pandey, R.; Fanello, S. R.; Wetzstein, G.; Zhu, J.; Theobalt, C.; Agrawala, M.; Shechtman, E.; Goldman, D. B.; Zollhöfer, M.: State of the Art on Neural Rendering. CoRR abs/2004.03805/, 2020, arXiv: 2004.03805, URL: https://arxiv.org/abs/2004.03805.

[Th16] Thies, J.; Zollhöfer, M.; Stamminger, M.; Theobalt, C.; Nießner, M.: Face2Face: Real-Time Face Capture and Reenactment of RGB Videos. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). S. 2387–2395, 2016.

[Th18] Thies, J.; Zollhöfer, M.; Theobalt, C.; Stamminger, M.; Nießner, M.: Headon: Real-Time Reenactment of Human Portrait Videos. ACM Trans. Graph. 37/4, Juli 2018, ISSN: 0730-0301, URL: https://doi.org/10.1145/3197517.3201350.

[Th21] Thiel, M.: „Deepfakes“ - Sehen heißt glauben? - Gefahren, gesetzgeberischer Handlungsbedarf, Konsequenzen für die biometrische Gesichtserkennung. Zeitschrift für Rechtspolitik/7, S. 202–205, 2021.

[Va17] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I.: Attention is All You Need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17, Curran Associates Inc., Long Beach, California, USA, S. 6000–6010, 2017, ISBN: 9781510860964.

[Za19] Zakharov, E.; Shysheya, A.; Burkov, E.; Lempitsky, V.: Few-Shot Adversarial Learning of Realistic Neural Talking Head Models, 2019, arXiv: 1905.08233.

[Zh22] Zhang, T.: Deepfake generation and detection, a survey. Multimedia Tools and Applications 81/, S. 6259–6276, 2022.

Informatik & Gesellschaft - Zukunft gestalten / IUG2022

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 683

Where bias can creep in

Gendersensibilität beim Einsatz von KI-Technologien im Kontext beruflicher Weiterbildung im Forschungsprojekt KIRA

Kerstin Raudonat[1], Monika Pröbster[1], Kai Schmieder[2], Simone Martinetz[2], Nicola Marsden[1]

Abstract: KI-Technologien können im Kontext beruflicher Weiterbildung den gezielten Kompetenzaufbau für künftige Berufsfelder unterstützen. Gleichwohl besteht die Gefahr, dass existierende Vergeschlechtlichungen im beruflichem Kompetenzaufbau reproduziert werden. Im Forschungsprojekt KIRA sollen berufliche Lernangebote KI-unterstützt auf individuelle Bedürfnisse zugeschnitten und passgenau vermittelt werden. Dabei soll schon während des Entwicklungsprozesses der Fortschreibung existierender Stereotypen entgegengewirkt werden. Hierzu werden mögliche Diskriminierungen in einem zirkulären Modell der Ko-Konstruktion konzeptionalisiert und so Ansatzpunkte für Reflexion und Intervention identifiziert. Im Beitrag werden vier Felder entlang dieses Kreislaufes möglicher Diskriminierung beleuchtet und es wird aufgezeigt, dass KI-gestützte Empfehlungen zur Kompetenzentwicklung nur mit einer ganzheitlichen Betrachtung wirklich fair sein können.

Keywords: Geschlechtergerechtigkeit; berufliche Weiterbildung; Kompetenzen; KI

1 Einleitung

Ethik und Gleichstellung spielen eine wesentliche Rolle beim Einsatz von künstlicher Intelligenz (KI): Gerade wenn es darum geht, auf Basis von KI Empfehlungen für grundlegende Lebensentscheidungen zu geben, sollte sichergestellt sein, dass sich keine Verzerrungen (Biases) auf Basis von Stereotypen in KI-Technologien materialisieren. Im Rahmen des Forschungsprojekts „KI-gestütztes Matching individueller und arbeitsmarktbezogener Anforderungen für die berufliche Weiterbildung“ (KIRA) sollen Empfehlungen für den Bereich der beruflichen Weiterbildung mittels KI ermöglicht werden. Es gilt, Gendersensibilität im Forschungs- und Entwicklungsprozess sicherzustellen. Dies betrifft einerseits das Thema Datensets, das für KI-Technologien von wesentlicher Bedeutung ist, und andererseits Themen aus dem Bereich beruflicher Weiterbildung wie beispielsweise vergeschlechtlichte Kompetenzzuschreibungen, Verzerrungen bei der Berufs- und Personalwahl sowie der Berufs- und Weiterbildungsberatung. Im Folgenden wird das Projekt KIRA skizziert und es

[1] Hochschule Heilbronn, Sozioinformatik, Max-Planck-Str. 39, 74081 Heilbronn vorname.nachname@hs-heilbronn.de

[2] Forschungs- und Innovationszentrum Kognitive Dienstleistungssysteme (KODIS), Fraunhofer IAO, Bildungscampus 9, 74076 Heilbronn vorname.nachname@iao.fraunhofer.de

doi:10.18420/inf2022_54

werden kritische Stellen im Forschungs- und Entwicklungsprozess identifiziert, an denen Verzerrungen ihren Weg in eine KI-basierte Lösung finden können (zur Bedeutung von Datensets und den Konsequenzen von Verzerrungen s. z.B. [BS16; Hu21]).

2 Das Projekt KIRA

Die digitale Transformation bringt nicht nur technologische Veränderungen mit sich, sondern fordert auch berufliche Weiterbildung und Qualifizierung von Mitarbeiter*innen für neue Tätigkeitsfelder. Im Verbundprojekt KIRA untersucht ein Konsortium an Forschungs- und Praxispartner*innen, wie die berufliche Weiterbildung datengestützt und mit künstlicher Intelligenz (KI) neu gestaltet werden kann. Dabei werden die datentechnischen Voraussetzungen für eine Individualisierung von beruflichen Lernangeboten mit Bezug auf unternehmens- bzw. arbeitsmarktpolitische Erfolgsmerkmale in den Fokus gerückt. Berufliche Lernangebote sollen auf individuelle Bedürfnisse zugeschnitten und passgenau vermittelt werden, so dass sie den individuellen Lernbedürfnissen entsprechen, gleichzeitig aber auch die arbeitsmarktpolitischen Anforderungen für den Erfolg von Lerninhalten berücksichtigen. Diese beziehen sich auf die konkreten Bedarfe von Unternehmen, aber auch auf übergeordnete Zielstellungen wie die Erschließung des Fachkräftepotenzials von Frauen für die Wirtschaft. Der Einbezug dieser übergeordneten Themen hat weitreichende Implikationen für die Einordnung und Gewichtung der individuellen Datenpunkte oder die Möglichkeit, aus vergangenen Daten zu extrapolieren. Gerade mit den in KIRA fokussierten neuen IT-bezogenen Qualifikationsbedarfen in transformationsintensiven Branchen und Unternehmen wird ein Bereich adressiert, der einerseits eine hohe Dynamik aufweist, andererseits von massiven Vergeschlechtlichungen in der Zuschreibung und in der Selbstwahrnehmung dieser Kompetenzen geprägt ist (vgl. Kapitel 3). Ziel der Modellbildung und des Matchings im Rahmen von KIRA ist es – nicht zuletzt im Sinne eines ethischen Einsatzes von KI – diesen Biases Rechnung zu tragen und so für die digitale Transformation der Arbeitswelt eine gleichberechtigte Teilhabe von Frauen in den zunehmend digitalisierten Arbeitsprozessen zu ermöglichen. Dies impliziert: KI-gestütztes Lernen und Qualifizieren muss fair sein.

3 Identifikation kritischer Stellen im Entwicklungsprozess

Im Projekt KIRA wird das Modell „Kreislauf der Diskriminierung“ [MRPht] verwendet, um frühzeitig – und auch im Verlauf - Faktoren im Entwicklungsprozess zu identifizieren, die kritisch im Hinblick auf Verzerrungen bezüglich Gender- und Diversitätsaspekten sind. Das Modell bildet den Zusammenhang zwischen gesellschaftlichen Biases und Diskriminierung in Anwendungen als einen wiederkehrenden Kreislauf der Produktion und Reproduktion sozialer Ungleichheiten im Kontext der Gestaltung von IT-Systemen ab. Abb. 1 visualisiert das zirkuläre Zusammenspiel von Produkt, Welt, Personen und Prozess, um eine differenzierte Betrachtung und Analyse von Diskriminierung und deren

Herstellung zu ermöglichen. Das Modell dient einerseits zur Erklärung von Diskriminierung und andererseits zum Identifizieren (segment)spezifischer Ansatzpunkte für Reflexion und Intervention.

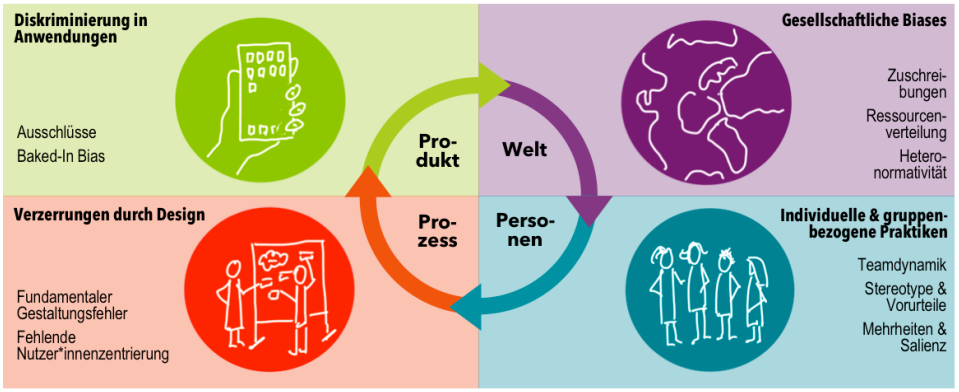


Abb. 1: Kreislauf der Diskriminierung [MRPht]

Im Folgenden werden anhand des Modells in den vier Feldern exemplarisch verschiedene kritische Faktoren beleuchtet, die speziell für die gendersensible Entwicklung von KI-Technologien im Kontext beruflicher Weiterbildung mit Fokus auf digitale Kompetenzen relevant sind. Auf dieser Basis werden Punkte identifiziert, an denen Interventionen für Geschlechtergerechtigkeit ansetzen können.

3.1 Produkt: Diskriminierung in Anwendungen

KI-Technologien sind nicht neutral und führen – wie sowohl wissenschaftlich als auch zunehmend öffentlich kritisiert – oftmals zu Verzerrungen und Diskriminierungen. Drei Ebenen, auf denen Biases auftreten können, können unterschieden werden: Biases in den Daten zum Training von KI-Anwendungen, Biases in den Algorithmen selbst und Biases im Kontext der Nutzung [Dr19; Me21].

Fehlende Diversität steht bei der Erhebung von Daten und Entwicklung von Datensets mehr und mehr in der Kritik. Unausgewogenes, fehlendes, korrumpiertes oder systematisch-verzerrtes Datenmaterial führt regelmäßig zur Diskriminierung von Mitgliedern gesellschaftlicher Gruppen wie z.B. Frauen und People of Color [BS16]. Unausgewogene Datensets zum Training von KI-Anwendungen führen dazu, dass Diskriminierungen via maschinellem Lernen in Anwendungen verankert werden - sie werden damit sozusagen unsichtbar und im Zuge nur bedingt erklärbarer Modelle eingeschränkt zugänglich. Zudem werden Datensets, die die Basis für maschinelles Lernen darstellen, häufig „used, shared, and reused with little visibility into the processes of deliberation that led to their creation“ [Hu21, S. 560]. Auch die Kategorienbildung (data labeling) kann solchen Verzerrungen unterworfen sein bzw. Verzerrungen einbringen oder fortsetzen. Somit ist es eine so essenzielle wie schwierige

Anforderung an den Einsatz von KI-Technologien, sicherzustellen, dass Diskriminierungen und Biases nicht in die Algorithmen der KI miteinfließen und infolgedessen diese Stereotype und Biases unbemerkt durch die KI repliziert werden.

Systeme, die KI-Technologien integrieren, werden zunehmend u.a. bei der Personalauswahl eingesetzt und bieten anschauliche Beispiele für Diskriminierung (s. Review bei [KW20]). So hat z.B. das von Amazon eingesetzte KI-System zur Beurteilung von internen Bewerber*innen Frauen systematisch herausgefiltert und somit auch keine Frauen zur Besetzung von internen Stellen vorgeschlagen. Diese „Empfehlung“ der KI-Software basierte auf dem zuvor trainierten Algorithmus, der die Einstellungen der vergangenen Jahre analysierte und zu dem Ergebnis kam, dass die Eigenschaft „Frau“ im Unternehmen keine Aussicht auf Erfolg hat [Da18]. Auch mit dem Versuch, Biases zu vermeiden, indem die Information zu bestimmten, potenziellen Quellen von Diskriminierung wie Geschlecht, Alter oder ethnische Zugehörigkeit als Variable aus dem Daten eliminiert wird, lässt sich das Problem nicht einfach lösen. Im Gegenteil: Diskriminierende Entscheidungsmechanismen können immer noch auftreten, sind aber schwerer zu entdecken [WBS18]. Teilweise führen auch andere Faktoren, die nichts mit der konkreten Datenbasis zu tun haben, zu unerwünschten Effekten [SG19]. Beispielsweise werden jungen Frauen bei Google seltener Werbeanzeigen für MINT-Stellen angezeigt, wenn der Algorithmus auf Kostenoptimierung ausgelegt ist, da Werbung für diese Zielgruppe vergleichsweise teuer ist [LT19].

3.2 Welt: Gesellschaftliche Biases

Das Feld Welt bezieht sich auf die Lebenswelten, in denen Personen leben und sozialisiert werden, die u.a. als Akteur*innen in Beratungs- und Entwicklungsprozessen zu berücksichtigen sind. Diese Lebenswelten sind geprägt von gesellschaftlichen Biases und sozialer Ungleichheit entlang unterschiedlicher Kategorien bzw. kategorialer Zuschreibungen. Eine wesentliche Kategorie ist dabei "Geschlecht". Stereotype und Vorurteile, die sich z.B. auf Geschlecht beziehen, beeinflussen Verhalten und können entsprechend auf aktionaler Ebene zu Diskriminierung von Personengruppen führen. Die moderne Stereotypenforschung zeigt, dass auch ohne die Existenz von Vorurteilen, nur auf Basis der Kenntnis von Stereotypen diskriminiert wird [KB21], also auch entgegen der Überzeugung einer Person.

Gerade im Hinblick auf MINT-Fächer und -Berufe und Kompetenzen, die im Zuge der Digitalisierung immer relevanter werden und als digitale Kompetenzen im Projekt KIRA im Fokus stehen, ergibt sich aus dem Zusammenwirken von Assoziationen ein Paket an Zuschreibungen und Abschreibungen, das sich sowohl im Kontext von Ausbildungs- und Berufswahl als auch in späteren Arbeitsfeldern und beruflicher Weiterbildung auswirkt. Noch immer gängige Assoziationen sind, Männlichkeit mit „technisch“ und „kompetent“ und Weiblichkeit mit „nicht-technisch“ und „sozial“ zu verknüpfen [El18; KW21]. Bereits Grundschulkinder schreiben Mathematik und mathematisch-logische Fähigkeiten eher Jungen und Männern als Mädchen und Frauen zu [Cv21]. Geschlechtsspezifische Fähigkeitszuschreibungen zeigen sich u.a. in vergeschlechtlichten Interessen [BLC17] und in den

MINT-bezogenen Fähigkeitsselbstkonzepten und Leistungsattributionen [ELP17]. Derart wirksame Geschlechterstereotype lassen sich auch bei Lehrkräften und Eltern nachweisen: Bei vergleichbaren Leistungen und Fähigkeiten in MINT-Fächern werden Jungen besser eingeschätzt als Mädchen [HS05]. Entsprechendes gilt für Männer und Frauen im Beruf [SE16], speziell bezogen auf Umgang mit digitalen Technologien [Fr20]. Tatsächlich zeigen Studien im Hinblick auf die Zuschreibung von Kompetenz generell, dass Frauen eher unterbewertet und Männer eher überbewertet werden [El18]. Dieser Unterschied in der Kompetenzwahrnehmung spiegelt sich auch auf Ebene der Selbsteinschätzung wieder – jedoch ohne entsprechendes geringeres Leistungskorrelat auf Seiten der Frauen; Männer hingegen überschätzen sich eher [Si03].

Im Bereich Berufswahl ist zudem der Einfluss von Eltern zu berücksichtigen: So zeigt z.B. die Langzeitstudie von Jacobs, Chhin und Bleeker [JCB06], dass die (u.a. geschlechtsstereotypen) Erwartungen von Eltern ein signifikanter Faktor sowohl im Hinblick auf die beruflichen Erwartungen der Kinder als auch deren Berufswahl sind. Bezüglich des Einschlagens mathematischer Karrieren zeigten Bleeker und Jacobs [BJ04] zudem, dass gerade die Erwartungen der Mütter im frühen Kindesalter einen deutlichen Einfluss auf späteres Wahlverhalten unabhängig von den tatsächlichen mathematischen Fähigkeiten der Kinder haben. Auch die Verknüpfung von Führungsrollen mit insbesondere dem männlichen Geschlecht [Gi17; Sc96] sind ein wichtiger Bias im Kontext beruflicher Weiterbildung, da Weiterqualifizierung häufig auf das zukünftige Übernehmen einer anderen Rolle – beispielsweise einer Führungsrolle – ausgerichtet ist.

Vor diesem Hintergrund ist im Projekt KIRA stets zu bedenken, dass bisherige Berufswege zumeist schon Biases aufweisen und entsprechend nicht unbedingt die individuellen Fähigkeiten und Interessen angemessen widerspiegeln. Eine Analyse von diesbezüglichen Dokumenten kann also immer nur ein Teil des KI-gestützten Matching-Prozesses im Kontext beruflicher Weiterbildung sein.

3.3 Personen: Individuelle und gruppenbezogene Praktiken

Die dargestellten Mechanismen auf Basis von Stereotypen finden ihren Niederschlag in individuellen und gruppenbezogenen Praktiken, die sich im Bereich beruflicher (Weiter)Bildung und im beruflichen Alltag zeigen, z.B. bei der Berufs- und Bildungsberatung, bei der Ausschreibung und Besetzung von Stellen in Unternehmen und in Teams, die Software für diese Einsatzbereiche entwickeln (z.B. KI-basierte Recruitment-Tools, siehe oben).

Im Feld der Berufsberatung tragen Geschlechterstereotype dazu bei, bestehende Ungleichheiten im Arbeitsmarkt zu perpetuieren. So zeigt beispielsweise die Studie von Ostendorf [Os05] zur Berufsberatung der Bundesagentur für Arbeit, dass diese der beruflichen Segmentation der Geschlechter nicht nur nicht entgegenwirkt, sondern diese mitverursacht. U.a. würden viele Berater*innen Mädchen nur dann gewerblich-technische Berufe anraten, wenn die Mädchen in Mathematik und Naturwissenschaften „besonders gute Schulnoten

vorweisen können sowie eine starke Persönlichkeit mitbringen, die sich in einer Männergesellschaft wird behaupten können“ [Os05, S. 126]. Auch im Rahmen von Beratungen eingesetzte Interessenstests weisen Gender-Biases auf (s. dazu z.B. [WH13]). Im Hinblick auf das Fach Informatik zeigt eine aktuelle Studie, dass – entgegen der häufigen Assoziation, dass gerade mathematisches Wissen ein guter Prädiktor für den Erwerb von Coding Skills sei – andere mögliche Prädiktoren diskutiert werden [RPJ17], z.B. sprachliche Begabung [Pr20]. In der Beratungspraxis scheint dies allerdings bisher wenig Beachtung zu finden, wobei gerade bei Schülerinnen, die sowohl in Mathematik als auch in Sprachen gut sind, zumeist eher die Sprachkompetenz wahrgenommen wird [SG18].

Zahlreiche Ansatzpunkte für das Wirken von Stereotypen finden sich bei Stellenausschreibungen, Stellenbesetzungen oder auch Vorstellungen und Plänen dazu, wer z.B. intern im Hinblick auf welche Position weiterqualifiziert wird. Bei Stellenbesetzungen werden bereits bei der Bewertung beruflicher Unterlagen Verzerrungen deutlich. Moss-Racusin et al. [Mo12] zeigten z.B., dass im Rahmen der Besetzung einer Stelle als Laborleitung identische Bewerbungsunterlagen unter dem Label „Mann“ als signifikant kompetenter und einstellungstauglicher eingestuft wurden als unter dem Label „Frau“. Zudem wurden ein höheres Einstiegsgehalt und mehr Karriere-Mentoring angeboten. Antragstellerinnen für ein Stipendium müssen deutlich höhere wissenschaftliche Leistung erbringen, um genauso kompetent wahrgenommen zu werden wie Antragsteller [WW08]. Krawczyk und Smyk [KS16] zeigten zudem, dass akademische Artikel von Autorinnen schlechter bewertet wurden als die von Autoren. Auch im Zuge der Gestaltung von Softwareprodukten für die Einsatzbereiche Bildung, Beratung und Personal-Recruitment werden geschlechtsbezogene Verzerrungen wirksam. Eine ausführliche Darstellung zu Biases und Diskriminierung bei der Entwicklung von IT-Systemen findet sich bei Marsden, Raudonat und Pröbster [MRPht].

3.4 Prozess: Verzerrungen durch Design

Für Gestaltungsprozesse gibt es unterschiedliche Methoden und Werkzeuge, die Entwickler*innen im Hinblick auf eine angemessene Konzeption und Berücksichtigung der Nutzer*innen unterstützen können bzw. sollen. Diese können durch ihre Konzeption oder in der Praxis ihrer Nutzung entweder hilfreich im Hinblick auf Gleichstellung sein oder wiederum Verzerrungen in den Prozess einbringen. Insbesondere der unsachgemäße oder unreflektierte Einsatz von Methoden und Tools birgt das Risiko, Stereotype in die Produktgestaltung zu integrieren, die in diskriminierenden Produkten resultieren. So führen beispielsweise die in der UX- und HCI-Praxis eingesetzten Empathy Maps häufig zu einer Explikation eigener Stereotype statt zu einer empathischen Konfiguration der Nutzer*innen [SD19]. Dies ist umso problematischer, da bei dem Einsatz von Tools und Methoden oftmals eine gewisse Standardisierung und Ergebnisqualität qua Tool/Methode unterstellt werden, die so aber nicht existieren.

Für den Bereich KI-Technologien sollen hier beispielhaft zwei Tools benannt werden, die zum Monitoring im Entwicklungsprozess von Anfang an eingesetzt werden können. So

kann zur Einschätzung von (geplanten) KI-Systemen die Kritikalitätsmatrix von Kraft und Zweig [Bu20; Zw22] genutzt werden. Entlang der Dimensionen Schadenspotenzial und Grad der Abhängigkeit werden hier KI-Systeme bewertet, die Entscheidungen unterstützen. Es werden also der konkrete Anwendungskontext und die damit verbundenen sozialen Prozesse berücksichtigt. Das zweite Tool zielt auf eine verantwortliche Gestaltung und Nutzung von Datensets: Das Framework for Dataset Development Transparency [Hu21] dient der Unterstützung von Entscheidungsfindung und Verantwortlichkeit entlang des Software Development Lifecycle. Es lenkt die Aufmerksamkeit unter anderem auf die häufig unbeachtete Arbeit und auf Entscheidungen, die zur Erstellung eines Datensets geführt haben - und macht damit auch an dieser Stelle Verantwortung adressierbar [Hu21].

4 Fazit und Ausblick

Die vorangegangenen Ausführungen zeigen anhand von verschiedenen Beispielen kritische Stellen im Hinblick auf Gendersensibilität und Biases im Kontext von KI-Technologien im Feld beruflicher Bildung auf, die im Projekt KIRA zu berücksichtigen sind. Ein vor diesem Hintergrund gendersensibler Einsatz von KI-Technologien in KIRA kann (und soll) die Chance bieten, Biases aktiv zu begegnen und die Erschließung digitaler Kompetenzen und digitalisierungsrelevanter Berufs- und Tätigkeitsfelder für z.B. Frauen zu fördern. Dies kann geschehen, indem z.B. Stereotype überwindende Empfehlungen gegeben werden, die an Interessen und Fähigkeiten orientiert sind, die aus der einfachen Betrachtung bisheriger Erwerbsbiographien so zunächst nicht offensichtlich erscheinen (z. B. Relevanz sprachlicher Begabung für Coding Skills, siehe oben). Auch die Rahmensetzung von Empfehlungen ist wichtig: Wenn zum Beispiel im Rahmen von Studien zu 'Future Skills' die Kompetenz 'Scrum' als eine technologische Fähigkeit kontextualisiert wird, ist davon auszugehen, dass eine Empfehlung, sich die Methode Scrum anzueignen, anders aufgenommen wird, als wenn diese eingebettet ist in den Kontext von Organisations- und Moderationsfähigkeiten.

Es besteht weiterer Forschungsbedarf, wie geeignete Interventionen im oben dargestellten Kreislauf der Diskriminierung im Zusammenhang mit Empfehlungen zum individuellen Aufbau digitaler Kompetenzen und digitalisierungsrelevanter Berufs- und Tätigkeitsfelder aussehen könnten. Wie dargestellt spielen Biases und Stereotype der beratenden Personen eine wesentliche Rolle und fließen in den Beratungsprozess ein – ein KI-gestütztes System läuft also Gefahr, vergeschlechtlichte Empfehlungen zu reproduzieren und so den Status quo zu reifizieren.

Wesentlich ist nicht zuletzt die Frage, wie es gelingen kann, dass gute Empfehlungen auch auf Akzeptanz stoßen bei den Personen, die sich weiterqualifizieren möchten. Schließlich ist die beste Empfehlung nur dann von Nutzen, wenn die beratene Person sie auch annehmen und für sich umsetzen kann. Selbststereotype können hier verhindern, dass den Empfehlungen gefolgt wird. Denn aufgrund der Allgegenwart von Stereotypen ist es – folgerichtig, aber paradoxerweise – so, dass im beruflichen Beratungsprozess genau jene Empfehlungen als

besonders plausibel und non-biased wahrgenommen werden, die gängigen Geschlechterstereotypen entsprechen [Wa21]. Interventionen in nur einem der vier Felder des Kreislaufes der Diskriminierung scheinen somit wenig erfolgsversprechend: Nur wenn alle vier Felder ganzheitlich betrachtet werden, können Empfehlungen zur Kompetenzentwicklung im Kontext von KIRA nachhaltig wirksam werden.

Literatur

[BJ04] Bleeker, M. M.; Jacobs, J. E.: Achievement in math and science: Do mothers' beliefs matter 12 years later? Journal of Educational Psychology 96/1, S. 97, 2004, ISSN: 1939-2176.

[BLC17] Bian, L.; Leslie, S.-J.; Cimpian, A.: Gender stereotypes about intellectual ability emerge early and influence children's interests. Science 355/6323, S. 389–391, 2017, ISSN: 0036-8075.

[BS16] Barocas, S.; Selbst, A. D.: Big data's disparate impact. California Law Review 104/, S. 671–732, 2016.

[Bu20] Bundesministerium für Arbeit und Soziales: Begleitband des Bundesministeriums für Arbeit und Soziales, 2020, URL: https://www.bmas.de/DE/Service/Publikationen/a811-eu2020-begleitband.html.

[Cv21] Cvencek, D.; Brečić, R.; Gaćeša, D.; Meltzoff, A. N.: Development of Math Attitudes and Math Self-Concepts: Gender Differences, Implicit–Explicit Dissociations, and Relations to Math Achievement. Child Development 92/5, e940–e956, 2021, ISSN: 0009-3920.

[Da18] Dastin, J.: Amazon scraps secret AI recruiting tool that showed bias against women. In: Ethics of Data and Analytics. Auerbach Publications, S. 296–299, 2018, ISBN: 1003278299.

[Dr19] Draude, C.; Klumbyte, G.; Lücking, P.; Treusch, P.: Situated algorithms: a sociotechnical systemic approach to bias. Online Information Review 44/2, S. 325–342, 2019, ISSN: 1468-4527.

[El18] Ellemers, N.: Gender stereotypes. Annual review of psychology 69/Jan, 69:275–298, 2018.

[ELP17] Ertl, B.; Luttenberger, S.; Paechter, M.: The Impact of Gender Stereotypes on the Self-Concept of Female Students in STEM Subjects with an Under-Representation of Females. Frontiers in Psychology 8/, 2017, ISSN: 1664-1078, URL: https://www.frontiersin.org/article/10.3389/fpsyg.2017.00703.

[Fr20] Franken, S.; Herr, V.; Kempf, L.; Mauritz, N.; Wattenberg, M.: Women Ressource 4.0 - Potenziale von qualifizierten Frauen, 2020, URL: https://www.fh-bielefeld.de/multimedia/Fachbereiche/Wirtschaft+und+Gesundheit/Forschung/Denkfabrik+Digitalisierte+Arbeitswelt/WomenRessource4_0_Abschlussbericht.pdf.

[Gi17] Gipson, A. N.; Pfaff, D. L.; Mendelsohn, D. B.; Catenacci, L. T.; Burke, W. W.: Women and Leadership. The Journal of Applied Behavioral Science 53/1, S. 32–65, 2017, ISSN: 0021-8863, URL: https://doi.org/10.1177/0021886316687247.

[HS05] Herbert, J.; Stipek, D.: The emergence of gender differences in children's perceptions of their academic competence. Journal of applied developmental Psychology 26/3, S. 276–295, 2005, ISSN: 0193-3973.

[Hu21] Hutchinson, B.; Smart, A.; Hanna, A.; Denton, E.; Greer, C.; Kjartansson, O.; Barnes, P.; Mitchell, M.: Towards accountability for machine learning datasets: Practices from software engineering and infrastructure. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency/, S. 560–575, 2021.

[JCB06] Jacobs, J. E.; Chhin, C. S.; Bleeker, M. M.: Enduring links: Parents' expectations and their young adult children's gender-typed occupational choices. Educational Research and Evaluation 12/4, S. 395–407, 2006, ISSN: 1380-3611.

[KB21] Kurdi, B.; Banaji, M. R.: Implicit social cognition: A brief (and gentle) introduction. Preprint on PsyArXiv/, 2021.

[KS16] Krawczyk, M.; Smyk, M.: Author s gender affects rating of academic articles: Evidence from an incentivized, deception-free laboratory experiment. European Economic Review 90/, S. 326–335, 2016, ISSN: 0014-2921.

[KW20] Köchling, A.; Wehner, M. C.: Discriminated by an algorithm: A systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. Business Research 13/3, S. 795–848, 2020, ISSN: 2198-2627.

[KW21] Kohlrausch, B.; Weber, L.: Gender Relations at the Digitalised Workplace: The Interrelation Between Digitalisation, Gender, and Work. Gender a výzkum/-Gender and Research 21/2, S. 13–31, 2021, ISSN: 2570-6578.

[LT19] Lambrecht, A.; Tucker, C.: Algorithmic bias? An empirical study of apparent gender-based discrimination in the display of STEM career ads. Management science 65/7, S. 2966–2981, 2019, ISSN: 0025-1909.

[Me21] Mehrabi, N.; Morstatter, F.; Saxena, N.; Lerman, K.; Galstyan, A.: A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR) 54/6, S. 1–35, 2021, ISSN: 0360-0300.

[Mo12] Moss-Racusin, C. A.; Dovidio, J. F.; Brescoll, V. L.; Graham, M. J.; Handelsman, J.: Science faculty's subtle gender biases favor male students. Proceedings of the National Academy of Sciences 109/41, S. 16474–16479, 2012, URL: http://www.pnas.org/content/109/41/16474.abstract.

[MRPht] Marsden, N.; Raudonat, K.; Pröbster, M.: Kreislauf der Diskriminierung - Gesellschaftliche Biases und Diskriminierung in IT-Systemen als wiederkehrender Kreislauf der (Re)Produktion sozialer Ungleichheiten./, eingereicht.

[Os05] Ostendorf, H.: Politische Steuerung durch Symbole und Verfahrensweisen—Die Mädchenpolitik der Berufsberatung. In (Henninger, A.; Ostendorf, H., Hrsg.): Die politische Steuerung des Geschlechterregimes. Springer, S. 115–137, 2005.

[Pr20] Prat, C. S.; Madhyastha, T. M.; Mottarella, M. J.; Kuo, C.-H.: Relating natural language aptitude to individual differences in learning programming languages. Scientific reports 10/1, S. 1–10, 2020, ISSN: 2045-2322.

[RPJ17] Román-González, M.; Pérez-González, J.-C.; Jiménez-Fernández, C.: Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. Computers in Human Behavior 72/, S. 678–691, 2017, ISSN: 0747-5632.

[Sc96] Schein, V. E.; Mueller, R.; Lituchy, T.; Liu, J.: Think manager—think male: A global phenomenon? Journal of organizational behavior 17/1, S. 33–41, 1996, ISSN: 0894-3796.

[SD19] Siegel, D.; Dray, S.: The map is not the territory: empathy in design. Interactions 26/2, S. 82–85, 2019, ISSN: 1072-5520.

[SE16] Steffens, M.; Ebert, I. D.: Frauen–Männer–Karrieren: Eine sozialpsychologische Perspektive auf Frauen in männlich geprägten Arbeitskontexten. Springer-Verlag, 2016, ISBN: 3658107502.

[SG18] Stoet, G.; Geary, D. C.: The gender-equality paradox in science, technology, engineering, and mathematics education. Psychological science 29/4, S. 581–593, 2018, ISSN: 0956-7976.

[SG19] Suresh, H.; Guttag, J. V.: A framework for understanding unintended consequences of machine learning. arXiv preprint arXiv:1901.10002 2/, 2019.

[Si03] Sieverding, M.: Frauen unterschätzen sich: Selbstbeurteilungs-Biases in einer simulierten Bewerbungssituation. Zeitschrift für Sozialpsychologie: ZFSP 34/3, S. 147–160, 2003, ISSN: 0044-3514.

[Wa21] Wang, C.; Wang, K.; Bian, A.; Islam, R.; Keya, K. N.; Foulde, J.; Pan, S.: Bias: Friend or Foe? User Acceptance of Gender Stereotypes in Automated Career Recommendations. arXiv preprint arXiv:2106.07112/, 2021.

[WBS18] Williams, B. A.; Brooks, C. F.; Shmargad, Y.: How algorithms discriminate based on data they lack: Challenges, solutions, and policy implications. Journal of Information Policy 8/1, S. 78–115, 2018, ISSN: 2381-5892.

[WH13] Wetzel, E.; Hell, B.: Gender-related differential item functioning in vocational interest measurement. Journal of Individual Differences 34/3, S. 170–183, 2013.

[WW08] Wenneras, C.; Wold, A.: Nepotism and sexism in peer-review. In (Wyer, M.; Barbercheck, M.; Giesman, D.; Özturk, H.; Wayne, M., Hrsg.): Women, Science, and Technology. Routledge, S. 64–70, 2008, ISBN: 0203895657.

[Zw22] Zweig, K. A.: Risk Matrix, 2022, URL: http://aalab.cs.uni-kl.de/resources/.

Digitalisierung – Brandbeschleuniger oder Helfer eines nachhaltigen Wandels?

Nachhaltigkeitskriterien für digitale Werkzeuge und Projekte

Karolin Eva Kappler[1] [2], Nicole Wolf[2], Tanja Kopp-Malek[2], Dominik Pataky[2], Christiane Bodammer[2], Virginia Holness[2] und Christoph Reithmair[2]

Abstract: Digitalisierung und Nachhaltigkeit sind zwei globale Transformationsprozesse, die wir – in Anlehnung an und in Nachfolge des Hauptgutachtens des Wissenschaftlichen Beirats der Bundesregierung Globale Umweltveränderungen (WGBU) „Unsere gemeinsame digitale Zukunft" – zusammen denken wollen. Dafür haben wir einen Kriterienkatalog für Privatpersonen, Unternehmen und politischen Akteuren entwickelt. Zwei dieser Kriterien sollen hier exemplarisch vorgestellt und diskutiert werden.

Keywords: Nachhaltigkeit, Digitalisierung, Software, Kriterien

1 Einleitung

Wir leben, arbeiten, konsumieren und kommunizieren immer digitaler – Musik wird gestreamt, Essen über eine Lieferplattform beim Restaurant nebenan bestellt und selbst die Partnersuche verläuft digital. Die Corona-Pandemie hat diese Entwicklung noch verstärkt, so dass dieser digitale Transformationsprozess unsere Gesellschaft und Wirtschaft fundamental prägt: Social Media, Big Data, Smart Devices, Internet of Things oder Blockchain sind nur einige Beispiele in diesem Kontext. Der zweite globale Transformationsprozess umfasst Nachhaltigkeit, da die Erde an ihre „planetarischen Leitplanken" [WG19a] kommt. Nachhaltigkeit im Digitalbereich ist jedoch ein bisher stark unterrepräsentiertes Thema, bei gleichzeitig intensivem Anwachsen von Digitalisierung in unterschiedlichen Arbeits- und Lebensbereichen. Aus diesem Grund haben wir uns zum Programm gemacht, Digitalisierung und Nachhaltigkeit zusammen zu denken und dabei persönliche, politische, ökonomische und institutionelle Aspekte mit zu berücksichtigen.

Warum ist das wichtig? Digitalisierung hat das Potenzial, einen wichtigen Beitrag zur Erreichung der Nachhaltigkeitsziele der Vereinten Nationen (SDG) zu leisten. So kann beispielsweise der Energie- und Ressourcenverbrauch reduziert oder die Mobilitätswende

[1] FernUniversität in Hagen, Fakultät für Wirtschaftswissenschaft, Lehrstuhl für Betriebswirtschaftslehre, insb. Betriebliche Anwendungssysteme, 58084 Hagen, karolin.kappler@fernuni-hagen.de
[2] Arbeitskreis Nachhaltigkeit, Fachbereich Informatik und Gesellschaft, Gesellschaft für Informatik, https://ak-nachhaltigkeit.gi.de/

doi:10.18420/inf2022_55

und die Entwicklung einer Kreislaufwirtschaft unterstützt werden. Das gilt es zu nutzen. Gleichzeitig birgt Digitalisierung jedoch die Gefahr, durch Effizienzorientierung, durch steigende Ressourcenverbräuche aufgrund von Reboundeffekten und Obsoleszenzen, durch Verknüpfung verschiedener Datenbanken und aus Machtinteresse Menschen zu kontrollieren, zu manipulieren und zu überwachen und damit Grund- und Menschenrechte (etwa den Schutz öffentlicher Räume und der Privatsphäre von Bürger:innen) zu unterminieren, aber auch Nutzer:innen zu überfordern.

Wir wollen deshalb Impulse setzen, wie Digitalisierung nicht nur Geschäfts- und Marktinteressen, sondern auch dem Gemeinwohl dienen kann. Darüber wollen wir nachdenken, neudenken, Gewohnheiten hinterfragen und konkrete Forderungen formulieren, diskutieren und umsetzen [Ko21].

2 Unser Verständnis von Nachhaltigkeit und Digitalisierung

2.1 Digitalisierung

Aus technischer Sicht stellt Digitalisierung die Konvertierung analoger Signale in eine digitale Form dar. Aus einer organisationstheoretischen und gesellschaftlichen Perspektive wird darunter die Übertragung von Aufgaben, die bisher vom Menschen übernommen wurden, auf den Computer verstanden. Der Wissenschaftliche Beirat der Bundesregierung Globale Umweltveränderungen (WGBU) definiert Digitalisierung als „die Entwicklung und Anwendung digitaler sowie digitalisierter Techniken, die sich mit allen anderen zivilisatorischen Techniken und Methoden verzahnt und diese erweitert“ [WG19a]. In diesem Sinne bringt Digitalisierung keine (neutralen) und rein technischen Systeme hervor, sondern sozio-technische Systeme, die soziale, ökonomische und ökologische Wechselwirkungen mit den Nutzer:innen und dem Umfeld hervorrufen.

Treibende Kraft hinter der Digitalisierung ist die Ökonomie und ihr Leitmotiv: das Gewinn- und Wachstumsstreben. Hinzu kommt auf der Ebene der Konsument:innen ein erhebliches Unterhaltungs- und Faszinationspotenzial, welches dem Trachten nach Anregung und Befriedigung des Geltungsdrangs des stets neugierigen und auf Selbstdarstellung erpichten Homo ludens entgegenkommt. „Nachhaltigkeit“ stellt aus dieser Sicht keinen Gewinn dar, sondern wird als Risiko verstanden.

Wie sähe jedoch eine Digitalisierung aus, die den Bedürfnissen der Nutzer:innen dient und nicht umgekehrt. Wie können Digitalisierungsprojekte so gestaltet werden, dass sie gemeinwohlorientiert der menschlichen Würde, der Teilhabe, demokratischen Prozessen und der Mündigkeit dienen und sich an den Prinzipien der Nachhaltigkeit ausrichtet [WG19a]?

2.2 Nachhaltigkeit

Neben der ökologischen Krise stellen sich folglich auch Fragen der sozialen Gerechtigkeit oder der politischen Teilhabe.

Der WBGU hat für die Navigation in diesem komplexen Gebiet einen „Normativen Kompass“ [WG19a] entwickelt, der die Einhaltung der Menschenwürde sicherstellen soll. Dies schließt mit ein und erfordert eine Balance zwischen dem Erhalt der natürlichen Lebensgrundlagen, d.h. die planetarischen Leitplanken müssen eingehalten und lokale Umweltprobleme vermieden werden, der Gewährleistung politischer Teilhabe, verstanden als universelle Mindeststandards für substanzielle, politische und ökonomische Teilhabe, sowie die Anerkennung von Vielfalt als zentrale menschliche Eigenart und als Bedingung für Wohlbefinden und Lebensqualität [WG19a].

Nachhaltigkeit besitzt folglich vier Dimensionen, die sich wechselseitig beeinflussen: (1) Ökologische Nachhaltigkeit: Planetare Grenzen, Umweltschutz, Ressourcen, …; (2) Ökonomische Nachhaltigkeit: Geschäftsmodelle, Lizenzen, Patente, …; (3) Soziale Nachhaltigkeit: Partizipation, Teilhabe, Arbeitsbedingungen, demokratische Praktiken, …; (4) Informationelle Nachhaltigkeit: Bildung, Information, Langzeitarchivierung, Öffentlichkeit, demokratische Prozesse…

„Nachhaltige Entwicklung heißt, Umweltgesichtspunkte gleichberechtigt mit sozialen und wirtschaftlichen Gesichtspunkten zu berücksichtigen. Zukunftsfähig wirtschaften bedeutet also: Wir müssen unseren Kindern und Enkelkindern ein intaktes ökologisches, soziales und ökonomisches Gefüge hinterlassen. Das eine ist ohne das andere nicht zu haben.“ [Ra11].

2.3 Nachhaltig digitale Möglichkeitsräume

Einerseits können digitale Lösungen helfen, ressourcenschonend zu wirtschaften und somit Umweltschäden beispielsweise durch Präzisionslandwirtschaft zu senken [WG19b], durch Vernetzung einen breiten Zugang zu Information zu gewährleisten oder Informationsasymmetrien abzubauen [Hi20]. Sie bergen jedoch auch die Gefahr von Reboundeffekten [Hi20, LPS20], geplanten (materiellen und psychologischen) Obsoleszenzen [JH22] sowie weiteren Risiken wie beispielsweise dem Verlust von Verfügungsgewalt über materielles Eigentum, das von Software abhängig ist [Hi20].

Nach welchen Prinzipien und Kriterien müssen folglich digitale Werkzeuge entwickelt werden, um eine digitale Nachhaltigkeitsgesellschaft anstatt einer Hyperkonsumgesellschaft zu unterstützen, Diversität anstatt Markt- und Machtmonopole zu fördern, eine Ungleichverteilung und die Durchsetzung zerstörerischer Einzelinteressen zu verhindern, digitale Mündigkeit, informierte (Eigen-)Verantwortlichkeit, Souveränität, Datenhoheit (d.h. ein offener, freier Zugang zu Gemeingütern) sowie Inklusion anstatt eines digitalisierten Überwachungsstaat zu ermöglichen?

3 Nachhaltigkeitskriterien für digitale Plattformen

Der Arbeitskreis Nachhaltigkeit in der Gesellschaft für Informatik hat sich zum Ziel gesetzt, einerseits die Themen Nachhaltigkeit und Digitalisierung zusammen zu bringen und als Mahner und Impulsgeber dafür Aufmerksamkeit zu schaffen. Andererseits sollen Kriterien (weiter-)entwickelt werden, um Entwickler:innen, Nutzer:innen, Bürger:innen und vielen mehr praktische Hilfestellungen zu geben, wie eine nachhaltige Digitalisierung aussehen könnte.

3.1 Erste Entwicklung

Eine erste Auswahl an Kriterien wurde vom Kollektiv „nachhaltig digital agieren“ entwickelt. Dafür hat das Kollektiv in einer Mindmap alle für die Gruppe relevanten Aspekte rund um das Thema „Plattformen“ nach den vier Nachhaltigkeitsdimensionen geordnet. Die Liste mit möglichen Kriterien war sehr lang, außerdem waren sie sehr unterschiedlich hinsichtlich ihrer Messbarkeit und Relevanz. Es fanden sich jedoch Kriterien, die an mehreren Stellen auftauchen und somit übergeordnet wichtig sein können. In einem nächsten Schritt wählten das Kollektiv gemeinsam die für sich relevantesten mit Hilfe einer internen Abstimmung aus. Bei der Reihenfolge und Gruppierung orientierte sich das Kollektiv an den drei Dimensionen des „Normativen Kompasses“ des WBGU. Daraus ist folgende Liste an Kriterien entstanden:

N – Natürliche Lebensgrundlagen	Kriterium 1: Angemessene Voreinstellungen Kriterium 2: Geringer Energieverbrauch Kriterium 3: Langlebigkeit – Update-Fähigkeit
T – Teilhabe	Kriterium 4: Barrierefreiheit Kriterium 5: Nutzungsautonomie Kriterium 6: Freie Software Kriterium 7: Plattformunabhängigkeit
E – Eigenart, Vielfalt, Wohlbefinden	Kriterium 8: Datensparsamkeit Kriterium 9: Minimales Tracking Kriterium 10: Angemessene Online-Werbung

Aus diesen Kriterien möchten wir hier exemplarisch zwei vorstellen (Auszug aus Ko21).

3.2 Forderung an die Industrie - Beispiel 1: Langlebigkeit – Update-Fähigkeit von Software (Kriterium 3)

Die Langlebigkeit von Gegenständen ist seit mehreren Jahrzehnten Teil der Diskussion über Umweltschutz und Ressourceneffizienz, hier sollen Aspekte zu der Langlebigkeit

von Software gesammelt werden. An einigen Stellen wird es Überschneidungen geben, da Software immer auch Hardware benötigt, um zu funktionieren.

Die Langlebigkeit von Software ist eine Sammlung von Eigenschaften, die für den nachhaltigen Betrieb einer Plattform wichtig sind. Dazu gehören die Quelloffenheit, die Update- bzw. Reparatur-Fähigkeit, die Portabilität und die Dokumentation.

Quelloffenheit: Jedes Computerprogramm liegt zunächst im Quelltext vor, der dann für eine entsprechende Hardware kompiliert wird, also in Maschinencode übersetzt wird. Jede Funktionsänderung oder Sicherheitsaktualisierung wird im Quelltext vorgenommen. Wenn die entwickelnde Organisation oder Person den Quelltext nicht veröffentlicht, sind die Betreiber:innen der Software darauf angewiesen, dass diese Änderungen von den Entwickler:innen vorgenommen werden. Insolvente Organisationen, berufliche Fluktuation und obsolete Hardware führen dazu, dass Software ohne vorliegenden Quelltext nicht oder nur unverhältnismäßig schwer aktualisiert werden kann. Der Quelltext muss in einer öffentlichen Form (nicht unbedingt kostenlos) vorliegen, damit der Betrieb der Software nachhaltig erfolgen kann.

Update-/Reparatur-/Support-Fähigkeit: Neben der notwendigen Bedingung der Quelloffenheit gibt es noch die hinreichende Bedingung, dass die Software entwickelnde Organisation oder Person die Aktualisierung auch erlaubt. Die Lizenz der Software verbietet einer Betreiber:in unter Umständen die eigenhändige Aktualisierung der Software, sondern verweist auf hauseigene App-Stores oder Update-Programme. Eine solche Abhängigkeit ist nicht nachhaltig. Die Update-Fähigkeit muss unabhängig von den Entwickler:innen bestehen.

Portabilität: Software ist in der Regel für eine bestimmte Hardware entwickelt. Eine Übertragung der Software auf eine neue Hardware kann technisch schwierig sein (in diesem Fall müsste man die Software mit Hilfe des öffentlichen Quelltextes neu kompilieren), sie kann aber auch schlicht unerwünscht sein. Lizenzen und Kopierschutzmodelle, die die Übertragung von einem System auf ein anderes verhindern, sind nicht nachhaltig. Die Portabilität muss unabhängig von den Hersteller:innen der Software gewährleistet sein.

Dokumentation: Warum eine Dokumentation wichtig ist, zeigt dieses komplett funktionstüchtige Programm: »:(){ :|:& };:«. Was tut es und warum? Das Programm startet rekursiv Kopien seiner selbst, verbraucht alle Systemressourcen und blockiert somit das System. Ohne die dazugehörige Dokumentation ist eine hinreichend komplexe Software nicht nachhaltig. Es gibt zwar Ansätze der automatischen Dokumentation von Software, aber es muss menschengeschriebene Erklärungen zur Funktionsweise von Software geben.

Community: Bei Open-Source-Projekten spielt die aktive Community eine entscheidende Rolle für den nachhaltigen Betrieb. Neben großen Flagschiffprojekten wie Ubuntu-Linux von Canonical oder der Firefox-Browser von Mozilla, die von großen Organisationen institutionell und finanziell unterstützt werden, gibt es zahlreiche kleine Projekte, die von

einer aktiven Community unterstützt werden, so dass eventuell auftretende Fehler behoben oder neue Funktionen ergänzt werden können.

Warum ist das für eine nachhaltige Entwicklung wichtig?

Neben den ökologischen Argumenten, die auch schon für den geringen Energieverbrauch gelten, ist dieses Kriterium vor allem für die informationelle Nachhaltigkeit entscheidend, also für die Erhaltung, den Zugang und die Weitergabe von Daten und Informationen unser menschliches Zusammenleben auf diesem Planeten betreffend. Es geht auch um die Vermeidung von Abhängigkeiten und so genannten Lock-In-Effekten, wenn der Betrieb einer Software von einer einzelnen Person oder Organisation abhängig ist.

Korrespondierende SDGs: Welche Nachhaltigkeitsziele werden durch dieses Kriterium adressiert / berücksichtigt / erfüllt?	
	9.1 Resiliente, zuverlässige und nachhaltige Infrastruktur entwickeln 9.4: Aktualisierung vorhandener Infrastrukturen, um sie Energie effizienter betreiben zu können
	12.4 + 12.5: Reduktion von Elektroschrott durch längere Nutzungsmöglichkeit 12.8: Informationen über ein harmonisches Zusammenleben für alle Leute bereitstellen
	13.2: Maßnahmen zur Klimakrise in nationale Unternehmensstrategien integrieren

3.3 Beispiel 2: Plattformunabhängigkeit (Kriterium 7)

Plattformen sind digitale, datenbasierte und algorithmisch strukturierende soziotechnische Infrastrukturen. Über Plattformen können Nutzer:innen Informationen austauschen, Arbeit und Märkte organisieren, Dienstleistungen anbieten, oder digitale sowie nicht-digitale Produkte vertreiben [Do20].

Plattformen wie etwa WECHANGE bestehen aus mehreren Komponenten (z.B. Forum, gemeinsame Dateiablage, Chat, Videokonferenz, Kalender, Etherpad), die aufeinander aufbauen. Das hier vorgeschlagene Kriterium soll deutlich machen, dass die Plattformbetreibenden die Abhängigkeiten zum einen erfassen und zum anderen so gestalten, dass die einzelnen Komponenten prinzipiell austauschbar sein sollen.

Eine wesentliche Abhängigkeit ist a) das Hosting und b) das eingesetzte Betriebssystem.

a) Das Hosting (deutsches Fremdwort ohne direkte Übersetzungsmöglichkeit; englisch: Internet Hosting Service) ist die Bereitstellung eines an das Internet angeschlossenen Servers (Host-Computers). Das Hosting erfolgt wahlweise als Dienstleistung durch einen so genannten Hoster oder die Plattformbetreibenden hosten es selbst.

b) Ein Betriebssystem ist eine notwendige Schnittstelle zwischen der Hardware und der einzusetzenden Software der Nutzer:in. Zu den bekanntesten Betriebssystemen auf Heimcomputern gehören die kommerziellen Betriebssysteme Microsoft Windows und Apple Mac OS sowie das freie Betriebssystem Linux, das von verschiedenen Hersteller:innen und Gruppen angeboten wird. Smartphones besitzen in der Regel entweder Apples IOS oder eine Android-Version (eine Linux-Variante, entwickelt von der Open Handset Alliance). Als Serverbetriebssystem kommen in der Regel Linux oder andere so genannte Unix-Varianten zum Einsatz. Bei allen Betriebssystemen ist für die Nachhaltigkeit die Update-Fähigkeit bzw. der Software-Lebenszyklus wichtig, also die Zeit, in der das Betriebssystem Updates erhält, um IT-Sicherheitslücken schließen zu können.

Unter Plattformunabhängigkeit verstehen wir also, dass eine Software auf verschiedenen Plattformen existieren kann (cross/multiplatform), während sie von keiner der Plattformen abhängig ist [BH06]. Plattformunabhängig sind also Softwareanwendungen, die auf vielen Betriebssystemen und Hardwareplattformen ohne Modifikation oder Anpassung ausgeführt werden können (wie z.B. das Schreiben von E-Mails).

Zusätzlich zu den obigen Definitionen wollen wir jedoch den Begriff weiter fassen und beziehen uns daher auf den zentralen Begriff der Pfadabhängigkeit. Die Pfadabhängigkeit bezeichnet in unserem Kontext die Situation, in der eine Technologie sich gegenüber einer anderen durchsetzt. Nicht unbedingt weil sie besser ist, sondern weil sie das Ergebnis vergangener Entscheidungen der Nutzer:innen sein kann [WG19a]. So ist es für große Software-Unternehmen nicht unüblich, Zwischenhändler von Geräten mit vorinstalliertem Betriebssystem dazu auffordern, keine alternativen Betriebssysteme auf ihren Geräten zu erlauben. Plattformunabhängigkeit heißt also auch, dass die Nutzer:innen die Softwarenutzung ohne erheblichen Aufwand zwischen verschiedenen Plattformen übertragen können.

Lizenzen und Kosten: Der Vertrieb eines Softwareprodukts als freie Software oder OSS ermöglicht mehr Menschen, sich in der Weiterentwicklung oder Problembehebung des Softwareprodukts zu beteiligen (siehe Kriterium „freie Software“). Die Softwarelizenz und die Kosten bei Supportanfragen entscheiden dabei mit, ob andere Entwickler:innen die Software für andere Plattformen unabhängig von den ursprünglichen Entwickler:innen weiter entwickeln (dürfen bzw. können) [HHM13].

Unterstütze Plattformen: Die Anzahl der von dem Softwareprodukt unterstützten Plattformen ist zwar wichtig, aber nicht alleine aussagekräftig für eine plattformunabhängige Software. Es spielt auch eine Rolle, wie weit verbreitet die unterstützten Plattformen sind (Windows als Betriebssystem ist zum Beispiel sehr weit verbreitet) und ob die Software die Plattformen gleichermaßen unterstützt [HHM13].

Wartbarkeit: Je mehr Plattformen eine Software unterstützt, umso mehr Aufwand benötigt ein Entwickler:innen-Team, Wartungsarbeit zu betreiben. Um die Wartbarkeit zu messen, helfen Metriken wie etwa die Aufzählung der Zeilen des Quelltextes, aus denen

das Programm besteht (engl.: Lines of Code, LOC), aber auch die Dokumentation, inkl. die der Modularität einer Software (siehe Kriterium „Nutzungsautonomie“) [HHM13].

Zugriff auf plattformspezifische Funktionen: Die Entwicklung einer plattformunabhängigen Software ist manchmal auf den kleinsten gemeinsamen Nenner der Funktionen beschränkt, die auf allen Plattformen verfügbar sind. Dennoch bietet der Zugriff auf eingebaute Funktionen der jeweiligen Plattformen, etwa Zugriff auf Kamera, Bildschirmleser (siehe Kriterium „Barrierefreiheit“), usw. eine bessere Nutzererfahrung [Da13].

Sicherheit: Eine Software, die verschiedene Plattformen unterstützt, muss dementsprechend gegen Schadsoftware (sog. Malware) der jeweiligen Plattformen robust sein: im schlimmsten Fall verbreitet eine solche Software Schaden über Plattformen hinaus [Li13]. Die Entwickler:innen müssen daher sowohl die plattformübergreifenden als auch plattformspezifischen Sicherheitslücken zeitnah erkennen und schließen können.

Warum ist das für eine nachhaltige Entwicklung wichtig?

Plattformunabhängigkeit ist ein wesentlicher Bestandteil für eine hochwertige Bildung (SDG 4), denn sie ermöglicht sowohl Lernenden als auch Lehrenden, über verschiedenen Plattformen hinaus Wissen zu erwerben bzw. zu ermitteln. Plattformunabhängiger Zugang zur Information hilft dann auch, weniger Ungleichheit (SDG 10) zu schaffen. Außerdem können Nutzer:innen ihre Plattformen länger anwenden, da sie eben nicht für jedes Softwareprodukt neue Plattformen brauchen. Diese führt dann zur verantwortungsvollen Konsum- und Produktionsmuster (SDG 12) (siehe Kriterium „Langlebigkeit und Update-Fähigkeit“).

Korrespondierende SDGs: Welche Nachhaltigkeitsziele werden durch dieses Kriterium adressiert / berücksichtigt / erfüllt?	
	4.7: Wissens und Kompetenzerwerb im Bereich Nachhaltigkeit für alle Wissbegierigen und Lernenden ermöglichen
	10.2: Förderung von sozialer, ökonomischer und politischer Inklusion
	12.8: Informationen über ein harmonisches Zusammenleben für alle Leute bereitstellen

4 Diskussion

Die einzelnen Kriterien sind nicht trennscharf, sie bedingen und beeinflussen sich wechselseitig. Die Nachhaltigkeitsziele der Vereinten Nationen, der hier in Auszügen

vorgestellte Kriterienkatalog und jeder andere Vorschlag zur schrittweisen Verbesserung der Lebenssituation auf diesem Planeten für alle Menschen weisen Zielkonflikte auf. Dies liegt in der Natur der Sache und ist Teil der gesellschaftlichen Auseinandersetzung. An einem konkreten Beispiel wollen wir das deutlich machen: Der Akkuverbrauch auf dem Gerät der Nutzer:in einer datenschutzfreundlichen, dezentralen Videokonferenz-Software ist derzeit in der Regel höher als der einer proprietären Lösung, die zentral die Inhalte bündeln kann. Die ökologische Forderung nach einem möglichst geringeren Energieverbrauch tritt in den Konflikt mit der sozialen Forderung nach Privatheit und Datenschutz. Es gibt keine mathematisch-logische Lösung dieses Konflikts, er muss ausdiskutiert und politisch entschieden werden. Dieser Katalog ist daher ausdrücklich nicht als Checkliste zu lesen, sondern als strukturierte Vorarbeit für die notwendige Diskussion, in welcher gemeinsamen digitalen Zukunft wir alle leben wollen.

5 Ausblick

Der Kriterienkatalog soll als ein kleiner Baustein dienen, um digitalen Umbrüche in Richtung Nachhaltigkeit auszurichten, denn nur so kann die Nachhaltigkeitstransformation gelingen. Digitalisierung droht ansonsten als Brandbeschleuniger von Wachstumsmustern zu wirken, die die planetarischen Leitplanken durchbrechen“ [WG19a].

Bei der Weiterentwicklung und Umsetzung der Kriterien treten Hürden auf, da einige Kriterien metrisch sind, andere wiederum nicht; manche Indikatoren sich mit ja oder nein beantworten, andere nicht. Darüber hinaus stellt sich die Frage, wo es Daten gibt, um entsprechende Aussagen treffen zu können? Und die Wirkungszusammenhänge zwischen den einzelnen Kriterien sowie fehlende Transparenz macht die Anwendung des Kriterienkatalogs sehr schnell sehr komplex. Aber ohne konkrete und praktische Unterstützung, die auch komplexe Entscheidungsprozesse transparent unterstützt, können Entwickler:innen und Nutzer:innen keine fundiert nachhaltigen Entscheidungen treffen.

6 Literaturverzeichnis

[BH06] Bishop, J.; Horspool, N.: Cross-Platform Development: Software that Lasts: Proceedings / 30th Annual IEEE/NASA Software Engineering Workshop. 24 - 28 April 2006, Columbia, Maryland. IEEE Computer Society, Los Alamitos, Calif., S. 119–122, 2006.

[Da13] Dalmasso, I. et al.: Survey, comparison and evaluation of cross platform mobile application development tools: 2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC 2013). Cagliari], Sardinia, Italy, 1 - 5 July 2013. IEEE, Piscataway, NJ, S. 323–328, 2013.

[Do20] Dolata, U.: Internet – Plattformen – Regulierung: Koordination von Märkten und Kuratierung von Sozialität. Working Paper, 2020.

[HHM13] Heitkötter, H.; Hanschke, S.; Majchrzak, T. A.: Evaluating Cross-Platform Development Approaches for Mobile Applications. In (Cordeiro, J.; Krempels, K.-H. Hrsg.): Web Information Systems and Technologies. 8th International Conference, WEBIST 2012, Porto, Portugal, April 18-21, 2012, Revised Selected Papers. Springer, Berlin, Heidelberg, S. 120–138, 2013.

[Hi20] Hilty, L. M.: Software und Nachhaltigkeit – Von der informationellen zur materiellen Selbstbestimmung. In (Göpel, M. et al. Hrsg.): Die Ökologie der digitalen Gesellschaft. S. Hirzel Verlag, Stuttgart, 2020.

[JH22] Jacobs, K.; Hörisch, J.: The importance of product lifetime labelling for purchase decisions: Strategic implications for corporate sustainability based on a conjoint analysis in Germany. Business Strategy and the Environment 4/31, S. 1275–1291, 2022.

[Ko21] Kollektiv nachhaltig digital agieren: Nachhaltigkeitskriterien für digitale Plattformen. Sammlung von Anforderungen für nachhaltigen Betrieb und Nutzung digitaler Plattformen und Werkzeuge mit Schwerpunkt Software. Ein Diskussionsbeitrag des Kollektivs nachhaltig digital agieren, Informatik 2021, 2021.

[Li13] Lindorfer, M. et al.: POSTER: Cross-platform malware. In (Sadeghi, A.-R. Hrsg.): Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security. ACM, New York, NY, S. 1425–1428, 2013.

[LPS20] Lange, S.; Pohl, J.; Santarius, T.: Digitalization and energy consumption. Does ICT reduce energy demand? Ecological Economics 176, 2020.

[Ra11] Rat für nachhaltige Entwicklung: Dialog der Verantwortung. Erwartungen des Nachhaltigkeitsrates an die Fortschreibung der nationalen Nachhaltigkeitsstrategie, Berlin, 2011.

[WG19a] WGBU - Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen: Transformation unserer Welt im Digitalen Zeitalter. Hauptgutachten, Berlin, 2019.

[WG19b] WGBU - Wissenschaftlicher Beirat der Bundesregierung Globale Umweltveränderungen: #SustainableDigitalAge – Illustriertes Fact Sheet. Factsheet, Berlin, 2019.

Das Universal Paradox

Im Spannungsfeld zwischen Universal Design und User Experience in einer Indoor-Navigations App für Menschen mit Einschränkungen

Andy Börner[1], Verena Traubinger[2], Philipp Stiens[1] und Maximilian Eibl[1]

Abstract: Barrierefreiheit und Inklusion stellen bei Webanwendungen häufig hohe Hürden dar. In der Technikentwicklung dient das Universal Design als Framework für ein Produkt für alle Nutzenden. In der Realität werden jedoch häufig einzelne Lösungen entwickelt. Anhand der Entwicklung einer Indoor-Navigations App werden drei Spannungsfelder in der Technikentwicklung aufgezeigt: Bekanntheit vs. Barrierefreiheit, Detailgrad vs. Überforderung und Individualisierung vs. Datensicherheit. Dabei zeigt sich, an welchen Stellen es Herausforderungen für eine gute User Experience zwischen individuellen Bedürfnissen und einer universellen Nutzbarkeit gibt.

Keywords: Universal Design; User Experience; Barrierefreiheit; Indoor Navigation; User-Centered

1 Einleitung

Zur Umsetzung von Barrierefreiheit bei Web-Anwendungen können Regelwerke wie etwa die WCAG 2.1 [WC22] genutzt werden. In der Technikentwicklung wird das Universal Design (UD) als Framework zur Umsetzung dieser Anforderungen genutzt [St01]. Während die Barrierefreiheit mögliche Barrieren erkennt und abbaut, ist das Ziel des UDs keine oder so wenig Barrieren wie möglich aufkommen zu lassen [Go09]. Das Gegenteil zu diesem holistischen Ansatz sind Insellösungen, die an spezifische Bedürfnisse angepasst sind. Als Beispiel sollen hier Navigationsapps dienen, die sich sowohl in Funktionalität als auch User Interface speziellen Anforderungen stellen. Viele Smartphone-Nutzende kennen Google Maps als Navigations-Anwendungen [GM22]. Einige Anwendungen decken jedoch spezifische Bedürfnisse von Anwendenden mit körperlichen und geistigen Einschränkungen ab: Wheelmap [WM22] zeigt die Barrierefreiheit von öffentlich zugänglichen Gebäuden an. Menschen mit Seheinschränkungen können hochkontrastige Kartenansichten [SD22] oder Indoor-Navigations Apps für einzelne Gebäude nutzen [LS22]. Das Universal Design beruht auf sieben Prinzipien [St01]. In diesem Paper sollen Paradoxe von sich widersprechenden UD Richtlinien in Bezug auf Barrierefreiheit exemplarisch an einer Indoor-Navigations App mit einem Fokus auf die Einbindung von Menschen mit Seh- und Geheinschränkungen [Ri20] gezeigt werden. Im Folgenden werden drei Bereiche vorgestellt, bei denen die Abwägung von Barrierefreiheit und den

[1] TU Chemnitz, Professur Medieninformatik, Straße der Nationen 62, 09111 Chemnitz, vorname.nachname@informatik.tu-chemnitz.de,
[2] TU Chemnitz, Professur Verteilte Systeme und selbstorganisierende Rechnersysteme, Straße der Nationen 62, 09111 Chemnitz, verena.traubinger@informatik.tu-chemnitz.de

doi:10.18420/inf2022_56

Prinzipien des UD das Design der App beeinflussten. Diese paradoxen Zwiespalte wurden nach dem Ansatz “Im Zweifel für die Einschränkung” aufgelöst.

2 Bekanntheit vs. Barrierefreiheit

Die meisten Interaktionen der Nutzenden finden in der App über die Kartendarstellung statt, die auch die Hauptoberfläche darstellt. Das UD fordert im Prinzip 4c die maximale Lesbarkeit von Informationen und in 1d, dass das Design ansprechend für alle ist [St01]. Im Kontext der App betrifft das also vorrangig die Kartendarstellung, die für alle lesbare Informationen enthält und ansprechend gestaltet sein soll.

Weltweit haben mindestens 2.2 Milliarden Menschen eine Form der Seheinschränkung [WH21]. Um ein UD zu schaffen, ist es deswegen wichtig die Karte so zu gestalten, dass so wenig Menschen wie nötig zur Nutzung eine Änderung in den Einstellungen vornehmen müssen. Im Rahmen des Projektes umfasst dies besonders Menschen mit Farbfehlsichtigkeiten (rot-grün, blau-gelb) und Menschen mit Seheinschränkungen, die nur sehr starke Kontraste deutlich wahrnehmen können.

Um Menschen mit und ohne Seheinschränkungen ein positives Nutzungserlebnis zu ermöglichen, werden Farben des Color Universal Design (CUD) verwendet [OI02]. Das CUD empfiehlt acht Farben, die auch für Menschen mit Farbfehlsichtigkeiten deutlich unterscheidbar sind. Es wurden sieben wichtige Kartenbereiche identifiziert, die eine Auswirkungen auf die Navigation und Orientierung von Menschen mit Einschränkungen haben.



Abb. 1: v. l. n. r. CUD im Simulator zur Farbfehlsichtigkeit, CUD, OSM-Farben im Simulator, OSM Farben - Obwohl keine Farben mehr vorhanden sind, ist es mit CUD im Vergleich zu OSM möglich, die einzelnen Elemente voneinander zu unterscheiden.

Jedem der sieben Bereiche wurde eine der acht Farben zugeordnet. Über die gewählten Farben wird eine Information über den jeweiligen Kartenbereich transportiert. Mögliche Gefahrenbereiche, z.B. durch verschiebbares Mobiliar, sind Rot (Vermillion) gekennzeichnet. Sanitärräume in Anlehnung an das dort zur Verfügung stehende Wasser Blau, Fenster in einem Himmel-Blau. Zusätzlich sind einige Bereiche wie Toiletten mit einem Symbol auf der Karte versehen, da sonst die Information rein über die Farbe übertragen werden würde. Trotz einer Verwendung von gewohnten Farben unterscheidet sich das erstellte Farbschema deutlich von Schemata der meist verbreitetsten Kartendiensten [GM22] [OS22], siehe Abbildung 1. Für Nutzende, die diese Kartendienste gewohnt sind, kann es zu Verwirrungen führen, wenn keine Legenden oder ein Tutorial zur Verfügung stehen.

Um im Sinne des UD allen Nutzenden eine User Experience nach ihrem Bedürfnis zu ermöglichen (Prinzip 2), besteht die Möglichkeit über ein Einstellungsmenü das Design der Karte zu ändern. Die Auswahl für die Kartenansicht besteht zwischen dem CUD-Farbschema und dem OSM-Standard-Schema.

3 Detailgrad vs. Überforderung

Die Kartografierung dient als Datengrundlage für das Rendering, das Routing und die späteren Navigationsanweisungen. Dabei wird auch die Grundlage für das Design bezüglich des 3. Prinzips des UD “Einfache und intuitive Bedienung” gelegt [St01]. Die Kartografierung beinhaltet die Informationsebenen der Gebäudeelemente, des Mobiliars und der Metainformationen (Raumnummern, Öffnungszeiten, …). Daraus können sogenannte Markante Punkte erstellt werden, die in der Navigation wichtige Orte wie Ein- oder Ausgänge, Toiletten oder Automaten für Nutzungszwecke kennzeichnen. Menschen mit Einschränkungen bevorzugen häufig zusätzliche multimodale Informationen zur Umgebung: Heizkörper und Fenster können durch veränderte Temperatur und Helligkeit Orientierungspunkte im Raum geben [Ts16], während große Räume oder die explizite Nennung fehlender Hilfsmittel Herausforderungen darstellen [En20].

Neben Informationen wie Materialien und Farben wird im Beispiel die manuelle bzw. automatische Öffnung von Türen kartografiert. Mit Ausnahme von Nutzgegenständen (Mülleimer, Steckdosen, etc.) wurde sich für eine übersichtlichere Darstellung (s. Abbildung 1) gegen die Kartografierung von Stühlen und anderem kleinen Mobiliar entschieden. Zusätzlich werden Flächen mit verrückbaren Gegenständen (Spiel- und Sitzecken) markiert, sowie Hilfsmittel zur Barrierefreiheit wie Handläufe, das Blindenleitsystem oder Audioausgaben. Die Kartografierung erfolgt nach dem OpenStreetMap Schema, das mit Tags arbeitet [OW22]. Damit ist es möglich, fehlende Hilfsmittel zu kennzeichnen und dies später bei den Routinginformationen anzugeben. Über die Einstellungen kann festgelegt werden, wie detailliert diese Informationen im Routing ausgegeben werden sollen, etwa indem nur die Navigation dargestellt wird oder auch Informationen wie ein Mülleimer auf der Route. In einer Onlinestudie wurde bereits die Relevanz verschiedener Navigationsinformationen für Menschen mit und ohne

Einschränkungen getestet [Tr21]. Dabei zeigt sich, dass eine Einteilung der Nutzenden in Gruppen keinen Mehrwert darstellt und jede Einstellung individuell möglich sein muss.

Wenn detailliertere Informationen ausgegeben werden sollen, muss beachtet werden, dass kein kognitiver Overload bei der Navigation stattfindet [Ts16]. Besonders Richtlinie 3 b des UDs "Ordnen Sie Informationen entsprechend ihrer Bedeutung an" [St01] ist dabei herausfordernd. Da die Relevanz der einzelnen Navigationsangaben sehr individuell ist, kann keine allgemeine Ordnung festgelegt werden. In Studien wurde vor allem von Menschen mit Einschränkungen die Notwendigkeit einer Planung zu barrierefreien Zugängen oder der allgemeinen Gebäudestruktur vor Antreten der Reise genannt [En20] [Fr21]. Diese erscheint, wenn die Indoor-Navigations App gestartet wird. Die Ausgabe des Routings kann aufgrund der umfassenden Kartografierung und mehrstufigen Auswahl in den App-Einstellungen einfach und flexibel vorgenommen werden.

4 Individualisierung vs. Datensicherheit

Um die App schon zu Beginn an die Bedürfnisse der Nutzen anzupassen, muss beim ersten Start der App der sogenannte Initialdialog durchlaufen werden. In diesem werden den Nutzenden einige Fragen gestellt, um daraus Informationen für die Darstellung des User Interfaces und der individuellen Navigationsinformationen zu erhalten. Prinzip 1b des UD weist darauf hin, dass die Segregation und Stigmatisierung von Nutzenden vermieden werden soll [St01]. Um dies zu erfüllen, werden von Nutzenden nicht konkrete Einschränkungen erfasst, sondern möglicherweise daraus folgende Präferenzen betrachtet. Jede Form der Befragung mit dem Ziel, die App an die Nutzenden anzupassen, führt zwangsweise dazu, dass Daten über sie gespeichert werden müssen. Diese Daten sind grundsätzlich gefährdet, da sie ein Ziel für Angreifende darstellen [Ha18]. Auch aus Datenschutzsicht ist es deshalb wünschenswert, so wenige Daten wie möglich zu speichern.

Nutzenden wird beim ersten Start der App die Möglichkeit gegeben, ihre Präferenzen hinsichtlich potenzieller Einschränkungen bei der Navigation und Nutzung zu individualisieren: ob Sie Aufzüge oder Treppen präferieren, breite Türen bevorzugt werden oder wie detailliert Informationen angezeigt werden sollen. Mit diesen Fragen können hilfreiche Informationen für die Navigation aufgenommen werden, ohne zu sehr in die Privatsphäre der Nutzenden einzudringen oder die Nutzenden in feste Gruppen (Geheingeschränkt, Seheingeschränkt) einzuteilen. Für das Routing durch das Gebäude ist es vollkommen unerheblich, aus welchen Gründen die Nutzenden entscheiden, dass sie z.B. Aufzüge präferieren. Das kann auch mit der Nutzung eines Kinderwagens oder schlicht der Bevorzugung von Aufzügen gegenüber Treppen zusammenhängen. Um den Initialdialog möglichst kurz zu halten, werden einige Daten direkt vom Smartphone erfasst, wie z.B. dunkle Themes oder Sprachausgabe.

Trotzdem werden die Daten zum Zwecke einer besseren User Experience gespeichert und sind damit grundsätzlich angreifbar. Angreifende können auch aus den vagen Informationen Rückschlüsse über die körperlichen Fähigkeiten der Nutzenden ziehen.

Zusätzlich stellt sich die Frage, ob die Nutzenden uns vertrauen die Daten nur für die Zwecke der App zu nutzen [WL17]. Eine Alternative bietet die Global Public Inclusive Infrastructure (GPII) [VTC13]. Ziel der GPII ist es, persönliche Einstellungen bezüglich der Barrierefreiheit sicher im Internet abzuspeichern und an beliebigen Geräten abzurufen. Dieser dezentrale Ansatz verspricht Vorteile in der Benutzung, da die Einstellungen an die App übertragen werden und somit z.B. der Initialdialog wegfallen könnte. Gleichzeitig birgt die Übertragung Gefahren, die bei lokaler Speicherung nicht entstehen.

Um die User Experience optimal an einzelne Personen anzupassen und Prinzip 2 des UDs zu erfüllen [St01], gibt es ein zusätzliches Einstellungsmenü mit Möglichkeiten um die Darstellung allgemein, die Übernahme der Smartphone Einstellungen und die Antworten auf die Fragen des Initialdialogs anzupassen. Durch die Vielzahl an Einstellungsmöglichkeiten besteht wiederum die Gefahr, dass Nutzende überfordert sind und im schlimmsten Fall für sie wichtige Einstellungen übersehen oder ignorieren.

5 Universal Design mit Einschränkungen

Die Grundsätze des UDs sind an 7 Prinzipien ausgerichtet [St01], die sich allerdings mehr an physischen Designs orientieren und nicht direkt auf Software-Entwicklungen angewandt werden können. Gleichzeitig sind diese Prinzipien bewusst unscharf und allgemein formuliert. Wie an den drei oben genannten Beispielen erkennbar ist, können komplexere Anforderungen zu einer Einschränkung der User Experience führen und damit gleichzeitig dem Prinzip der einfachen und intuitiven Bedienung widersprechen [St01]. Neben einer Anwendung für die größtmögliche Basis an Nutzenden, wird das UD nach [IS03] noch weiter gefasst: „Universal design denotes a process more than a definite result." Ziel dieses Papers ist es zu zeigen, wie ein UD für eine Indoor-Navigations App holistisch umgesetzt werden kann. Obwohl die Entscheidungen erst noch getestet werden müssen, wurden durch das Anwenden des UD einige mögliche Problemstellen gefunden, die in Testungen weiter betrachtet werden. Musste sich bei den hier vorgestellten Beispielen zwischen den Bedürfnissen von Menschen ohne Einschränkung und Menschen mit Seh- oder Geheinschränkungen entschieden werden, wurden letztere präferiert. Zum einen soll vermieden werden, dass Menschen mit Einschränkungen aufgrund von ästhetischen oder funktionalen Gewohnheiten weiteren Barrieren ausgesetzt sind. Zum anderen bietet eine größere Barrierefreiheit mehr Vorteile für alle Nutzenden. Für diese Mehrwerte lässt sich bei Abwägungen im Design also sagen: Im Zweifel für die Einschränkung!

Acknowledgement

Diese Einreichung wurde durch den Europäischen Sozialfonds (ESF) mit der Projektnummer 100382183 finanziert.

Literaturverzeichnis

[En20] Engel, C. et.al.: Travelling more independently: A Requirements Analysis for Accessible Journeys to Unknown Buildings for People with Visual Impairments, The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS ’20), Greece, S. 1-11, 2020.

[Fr21] Franzkowiak, L. et.al.: A blind spot in indoor navigation - needs and requirements of visually impaired people, 12th Media Psychology Conference (MediaPsych 2021), Aachen, 2021.

[Go09] Gossett, A.; et.al.: Beyond access: A case study on the intersection between accessibility, sustainability, and universal design. Disability and Rehabilitation: Assistive Technology 4/09, S. 439-450, 2009.

[GM22] GM, Google Maps, https://www.google.de/maps, Stand: 12.05.2022.

[Ha18] Hamidi, F. et.al.: Who Should Have Access to my Pointing Data?: Privacy Tradeoffs of Adaptive Assistive Technologies. ASSETS '18: Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility, Galway Ireland, S. 203-216, 2018.

[IS03] Iwarsson, S.; Ståhl, A.: Accessibility, usability and universal design—positioning and definition of concepts describing person-environment relationships. Disability and Rehabilitation 25/09, S. 57-66, 2003. [LS22] LS, Loud Steps Indoor navigation for the visually impaired, http://loudsteps.com/, Stand: 12.05.2022.

[OI02] Okabe, M.; Ito, K.: Color Universal Design (CUD)- How to make figures and presentations that are friendly to Colorblind people -, https://jfly.uni-koeln.de/color/, Stand: 11.05.2022.

[OS22] OS, OpenStreetMap, https://www.openstreetmap.org, Stand: 08.07.2022.

[OW22] OW, OpenStreetMap Wiki, https://wiki.openstreetmap.org/wiki/Tags, Stand: 06.07.2022.

[Ri20] Richter, J. et.al.: Dynamic indoor navigation and orientation system for people with impairments. In (Alt, F.; Schneegass, S.; Hornecker, E. Hrsg.): Mensch Computer 2020, New York, 473-477, 2020.

[SD22] SD, Stamen Design Stamen Toner Map, http://maps.stamen.com/toner/#12/37.7706/-122.3782, Stand: 12.05.2022.

[St01] Story, M.F.: Principles of universal design. Universal design handbook, Second Edition, McGraw-Hill, 2001.

[Tr21] Traubinger, V. et.al.: The Right Data at the Right Moment for the Right Person — User Requirements and Their Implications for the Design of Indoor Navigation Systems. 2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN), Lloret de Mar Spain, S. 1-8, 2021.

[Ts16] Tscharn, R. et.al.: Turn Left After the Heater‘: Landmark Navigation for Visually Impaired Users. Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, Reno Nevada USA, S. 295-296, 2016.

[VTC13] Vanderheiden, G.; Treviranus, J.; Chourasia, A.O.: Creating a Global Public Inclusive Infrastructure, 15th Media International ACM SIGACCESS Conference on Computers and Accessibility (Association for Computing Machinery), Bellevue Washington, 2013.

[WC22] WC, Web Content Accessibility Guidelines (WCAG) 2.1, https://www.w3.org/TR/WCAG21/, Stand: 06.07.2022.

[WH21] WHO, World Health Organization Blindness and vision impairment, https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment, Stand: 11.05.2022.

[WL17] Wang, E.; Lin, R.: Perceived quality factors of location-based apps on trust, perceived privacy risk, and continuous usage intention. Behaviour & Information Technology, 01/17, S. 2-10, 2017.

[WM22] WM, Wheelmap, https://wheelmap.org/, Stand: 12.05.2022.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 711

Die ethische Position von Informatikstudierenden bezüglich der Entwicklung von Informatiksystemen

Eine Untersuchung des Effekts der Teilnahme an einer Lehrveranstaltung aus dem Bereich „Informatik und Gesellschaft“

Nils Pancratz[1], Matthias Matzner[2]

Abstract: Im Rahmen ihrer (universitären) Ausbildung müssen Informatikerinnen und Informatiker unumstritten auch für ethische Aspekte im Zusammenhang mit der Entwicklung, Gestaltung, Nutzung und Wirkung von Informatiksystemen sensibilisiert werden. Im Informatikstudium finden ethische Fragestellungen üblicherweise Einzug in die Lehrinhalte von Veranstaltungen, die versuchen, den Inhaltsbereich Informatik und Gesellschaft möglichst multidimensional (gesellschaftlich, ökonomisch, rechtlich, politisch, ethisch, ...) abzudecken. Der Beitrag geht der Frage nach, inwiefern die Teilnahme an einer entsprechenden Lehrveranstaltung eine Auswirkung auf die ethische Position der Studierenden hat. Die Ergebnisse deuten darauf hin, dass die Teilnahme an der untersuchten Veranstaltung die ethische Position kaum beeinflusst, was mit Blick auf die Adäquatheit der Adressierung ethischer Kompetenzen in entsprechend vielschichtigen Veranstaltungen diskutiert wird.

Keywords: Ethik, Ethische Position, Informatik und Gesellschaft, Hochschuldidaktik

1 Einleitung, Motivation und Fragestellung

Aus der wachsenden Bedeutung ethischer Aspekte bei der Entwicklung und Gestaltung von Informatiksystemen und insbesondere auch aus der im Zusammenhang mit aktuellen Trendthemen wie Künstlicher Intelligenz oder Big Data in breiten Teilen der Gesellschaft geführten Diskussionen zu gesellschaftlichen Auswirkungen von Informationstechnologien erwachsen entsprechende Erwartungen und Anforderungen an die Ausbildung einer *ethischen Position* von Informatikerinnen und Informatikern. Im Rahmen der Ausbildung von Informatikstudierenden werden ethische Aspekte an deutschen Hochschulen überwiegend in Lehrveranstaltungen zu dem Gebiet „Informatik und Gesellschaft“ (IuG) behandelt, wenngleich sie dort meist nur einen untergeordneten Raum einnehmen. Auch in den Studienordnungen der verschiedenen Informatikstudiengänge an der Universität Oldenburg finden sich ethische Fragestellungen schwerpunktmäßig in den Lehrinhalten einer Wahlpflichtveranstaltung, die den entsprechenden Inhaltsbereich möglichst multidimensional (gesellschaftlich, ökonomisch, rechtlich, politisch, ethisch, ...) abzudecken versuchen soll. Aus dieser für das Informatikstudium an deutschen Hochschulen typischen Aus-

[1] Stiftung Universität Hildesheim, Institut für Mathematik und Ang. Informatik, pancratz@uni-hildesheim.de
[2] Carl von Ossietzky Universität Oldenburg, Department für Informatik, matthias.matzner@uol.de

doi:10.18420/inf2022_57

gangslage leitet sich für die vorliegende Untersuchung folgende Forschungsfrage ab: *Inwiefern verändert die Teilnahme an der Lehrveranstaltung „Informatik und Gesellschaft" die ethische Position bezüglich der Entwicklung von Informatiksystemen bei Informatikstudierenden der Universität Oldenburg?*

2 Theoretischer und methodischer Hintergrund der Untersuchung

2.1 Zur ethischen Position und ihrer Rolle im Informatikstudium

Ethik bezeichnet gemeinhin die praktische Wissenschaft von der *Moral*, also Normensystemen für das Verhalten von Menschen; *ethisch* bedeutet demnach, dass etwas in den Gegenstandsbereich der Ethik fällt oder diesem Wissensgebiet zugehörig ist [Hü21]. Nach Ansicht von Capurro soll Ethik per Definition Menschen in ihren moralischen Entscheidungen helfen [Ca17]. Mit Blick auf Informatikerinnen und Informatiker unterschiedlichster Fachrichtungen und Tätigkeitsfelder ergeben sich durch die möglichen Folgen ihres Wirkens Anforderungen an entsprechend professionelle ethische Kompetenzen, ein ethisches Grundverständnis sowie eine grundsätzliche ethische Sensibilität. Informatik-Fachverbände wie bspw. die *Gesellschaft für Informatik* (GI) oder die *Association of Computing Machinery* (ACM) begannen daher bereits vor Jahrzehnten mit der Formulierung (und regelmäßigen Überarbeitung) von Kodizes, die Informatikerinnen und Informatikern eine Hilfestellung bei Entscheidungen bieten sollen, die ethische Aspekte beinhalten. Solche ethischen Aspekte betreffen insbesondere *Regularien* zur Einschränkung in der Entwicklung von Systemen bspw. im Falle möglicher Gefahren durch diese, die *Privatsphäre* der Nutzerinnen und Nutzer der Systeme, das *Wissen* über (einheitliche) ethische Rahmenbedingungen im Zusammenhang mit der Entwicklung der Systeme, die *gesellschaftliche Verantwortung*, die mit möglichen Auswirkungen von technologischen Entwicklungen für die Gesellschaft einhergeht und die *organisationale Verantwortung*, die für Unternehmen erwächst [BSM18]. Diese vielschichtigen ethikbezogenen Aspekte werden in diesem Beitrag als *ethische Position* zusammengefasst. Brandenburg u. a., deren Fragebogen *Epos* in dieser Untersuchung Verwendung findet, verstehen bspw. „die Einstellung von Menschen zu ethischen Aspekten in der Softwareentwicklung [...] als *ethische Position*" [BSM18][3].

Aus der Bedeutung einer ethischen Position bei Informatikerinnen und Informatikern erwachsen entsprechende Anforderungen an die Informatikausbildung an Hochschulen. In den Studienordnungen zu verschiedenen Informatikstudiengängen in Deutschland gibt es derzeit jedoch kaum Module, die sich ausschließlich mit ethischen Fragestellungen beschäftigen [GSM21] und entsprechend ausschließlich auf die Ausbildung einer ethischen Position bei den Studierenden ausgelegt sind. Im europäischen Vergleich scheint ethischen Aspekten im Informatikstudium in Deutschland überdies eine unterdurchschnittli-

[3] *Regularien*, *Privatsphäre*, *Wissen*, *gesellschaftliche Verantwortung* und *organisationale Verantwortung* sind Subskalen von *Epos* [BSM18].

che Relevanz zuzukommen [St21]. Entsprechende Kompetenzen würden vielmehr überwiegend in Modulen adressiert, die sich neben ethischen Fragestellungen auch mit anderen, interdisziplinären Themen, insbesondere rechtlichen Aspekten und Fragen der Selbst- und Arbeitsorganisation, befassen und am häufigsten mit „Informatik und Gesellschaft" benannt werden [GSM21]. Lehrveranstaltungen zu dem Bereich IuG fokussieren dabei üblicherweise neben ethischen Aspekten der Nutzung und Gestaltung von Informatiksystemen auch auf gesellschaftliche, ökonomische, rechtliche und politische Kontexte [PPH13]. Es erscheint somit naheliegend, dass die jeweiligen Modulverantwortlichen sich an dem entsprechenden Inhaltsbereich der *Empfehlungen für Bachelor- und Masterprogramme im Studienfach Informatik an Hochschulen* der GI orientieren [GSM21].

Auch an der Universität Oldenburg, die mit *Informatik*, *Wirtschaftsinformatik und Engineering of Socio-Technical Systems* sowie verschiedenen *Lehramtsstudiengängen* mehrere Informatikstudiengänge zu ihrem Studienangebot zählt, ist die Thematisierung ethischer Aspekte der Informatik im Rahmen der entsprechenden Studienordnungen überwiegend in einer Wahlpflichtveranstaltung mit dem Titel „Informatik und Gesellschaft" (4 SWS) vorgesehen. Lehramtstudierende *können* das zugehörige Modul (6 ECTS) im Master, die Studierenden der übrigen Informatikstudiengänge im Bachelor belegen[4].

2.2 Verwandte Arbeiten und Ableitung von Hypothesen

Zuletzt war international zwar ein zunehmendes Bestreben zu vernehmen, ethische Inhalte in die Informatikausbildung an Hochschulen zu integrieren, jedoch liegen bislang nur wenige Arbeiten vor, die die Wirksamkeit entsprechender Initiativen evaluieren [Ho22]. Horten u. a. bspw. untersuchten die Auswirkungen der Teilnahme an ihrer entsprechenden Veranstaltung und bezeichnen ihre Arbeit dabei als die erste quantitative Untersuchung der Integration ethischer Inhalte in die Informatikausbildung. Sie kommen zu dem Ergebnis, dass sowohl das *Interesse* an ethischen Aspekten in der Informatik als auch die *Selbstwirksamkeitserwartung* mit Blick auf ethische Kompetenz durch die Teilnahme steigen [Ho22]. Hieraus lässt sich die Hypothese ableiten, dass auch die *ethische Position* von Studierenden nach der Teilnahme an einer entsprechenden Lehrveranstaltung stärker ausgeprägt ist als davor.

Lehrende auf dem Gebiet IuG berichten davon, dass der Frauenanteil in der Forschung in dem entsprechenden Gebiet „deutlich höher [ist] als in anderen Teilgebieten der Informatik" [PPH13]. Hieraus leiten bspw. Paech u. a. die Vermutung ab, „dass sich Frauen mehr für ein Informatik-Studium mit explizitem Anteil sozioinformatischer Inhalte interessieren würden" [PPH13]. Es erscheint daher mit Blick auf die Forschungsfrage einer Untersuchung würdig, inwiefern sich die Teilnahme an einer Wahlpflichtveranstaltung zu IuG auf

[4] Überdies können auch Bachelorstudierende informatikferner Studiengänge wie bspw. Wirtschaftswissenschaften oder Nachhaltigkeitsökonomik im Rahmen ihrer überfachlichen Qualifikation im Professionalisierungsbereich eine Leistung im Modul „Informatik und Gesellschaft" erbringen.

die Veränderung der ethischen Position bei weiblichen Studierenden anders auswirkt als bei ihren männlichen Kommilitonen.

2.3 Durchführung der Untersuchung und Kurzdarstellung der Ergebnisse

Zu Beginn und Ende des Wintersemesters 2020/2021 wurde an der Universität Oldenburg in einem Pre-Post-Design die *ethische Position*[5] der an der Veranstaltung „Informatik und Gesellschaft“ teilnehmenden Studierenden (sowie demografische Variablen wie Alter, Studiengang und Geschlecht) erhoben. Zur ersten Erhebung wurden Daten von 109 Studierenden erhoben, von denen 66 auch den zweiten Messmoment komplettierten. Unter den Studierenden, die an beiden Untersuchungen teilnahmen, befanden sich 58 männliche und 8 weibliche Studierende.

Zur Überprüfung von eventuellen Unterschieden zwischen der Pre- und Post-Messung sowie evtl. erklärenden Faktoren dieser Unterschiede wurden Inferenzialprüfungen mittels Varianzanalyse mit Messwiederholung erstellt (vgl. Tab. 1). Es konnten für keine der Subskalen aus *Epos* statistisch signifikante Veränderungen von der Pre- zur Post-Messung festgestellt werden. Lediglich für die Subskala *Wissen* wurde ein Trend in Richtung statistischer Signifikanz für einen Interaktionseffekt mit Geschlecht festgestellt und mit paarweisen Vergleichen näher verortet: Weibliche Studierende bewerten die Relevanz, dass Informatikerinnen und Informatiker über Wissen zu ethischen Rahmenbedingungen im Zusammenhang mit der Entwicklung von Informatiksystemen verfügen sollten, nach der Teilnahme an der Lehrveranstaltung höher als davor [0,04; 1,04]. Bei männlichen Studierenden trat keine eindeutige Änderung in der *Epos*-Subskala *Wissen* auf [-0,24; 0,14].

	Pre-Post		Pre-Post* Geschlecht		Pre-Post* Alter	
	F	p	F	p	F	p
I. Regularien	0,03	0,874	2,32	0,133	0,00	0,966
II. Privatsphäre	0,19	0,663	0,49	0,489	0,10	0,759
III. Wissen	0,94	0,337	0,83	0,070	1,60	0,210
IV. Gesellschaftliche Verantwortung	0,01	0,927	1,58	0,213	0,05	0,817
V. Organisationale Verantwortung	0,37	0,545	0,53	0,471	0,33	0,570

Tab. 1: Varianzanalyse mit Messwiederholung: Innersubjektfaktor Pre-Post, Innersubjektfaktor Geschlecht und Kovariate Alter. Alle Levene’s Tests F(1; 61).

[5] Hierzu kam das validierte Fragebogeninstrument *Epos* [BSM18] zum Einsatz, wobei der darin verwendete Begriff „Software“ für die vorliegende Untersuchung zu „Informatiksystem“ erweitert wurde.

3 Diskussion und Zusammenfassung der Ergebnisse

Die Forschungsfrage kann nur teils mit positivem Befund beantwortet werden: Die Ergebnisse zeigen bei Studentinnen einen Trend zur positiven *Epos*-Veränderung in der Subskala *Wissen*, jedoch konnte keine Veränderung der ethischen Position im Allgemeinen festgestellt werden. Angesichts der hohen Relevanz einer Sensibilisierung für ethische Aspekte bei der Entwicklung und Gestaltung von Informatiksystemen scheint dieses Ergebnis entsprechende Desiderate für Bildung und Forschung aufzudecken: So wirft die vorliegende Untersuchung der Veränderung der ethischen Position durch die Teilnahme an einer Lehrveranstaltung insbesondere die Frage auf, ob das didaktische Format (von der inhaltlichen Vielschichtigkeit über die Interaktionsgestaltung bis zur Leistungsüberprüfung) für die Entwicklung einer ethischen Position adäquat ist. Evtl. ist ein stark auf selbstgesteuertes Lernen ausgelegtes Seminar alleine stehend nur bedingt geeignet, ethische Fragestellungen zu er- und zu bearbeiten und dadurch implizit eine ethische Position auszubilden. In weiteren Durchläufen wäre daher neben einer stärkeren Schwerpunktsetzung auf ethische Aspekte in der Lehrveranstaltung auch der Effekt eines stärker inputorientierten Vorgehens bspw. durch ergänzende Vorlesungsinhalte zu untersuchen.

3.1 Selbstkritische Auseinandersetzung mit der Aussagekraft der Ergebnisse

Die vorliegende Untersuchung unterliegt dem Forschungsdesign geschuldet einigen Einschränkungen: Grundsätzlich erscheint es bedeutsam, kritisch zu hinterfragen, ob eine Selbstauskunft die implizite ethische Position angemessen erfasst. Insbesondere im Zusammenhang mit der Erhebung der ethischen Position in einer Lehrveranstaltung muss überdies mit Effekten sozialer Erwünschtheit gerechnet werden, die Frageitems können schließlich den Anschein von Examinierungsfragen erwecken (der Erstautor war z. Zt. der Datenerhebungen modulverantwortlich). Überdies wurden externe Störfaktoren nicht berücksichtigt: Besonders die Einschränkungen der möglichen Interaktionen zwischen Studierenden im Zusammenhang mit den Kontaktbeschränkungen zur Pandemiebekämpfung im zweiten nahezu vollständig digitalen Semester könnten Effekte der Teilnahme an einer entsprechenden Lehrveranstaltung negativ beeinflusst haben. Außerdem ist die Repräsentativität eingeschränkt, weil die Studierenden schließlich allesamt an einer bestimmten Lehrveranstaltung an einem bestimmten Hochschulort teilgenommen haben.

3.2 Fazit und Ausblick

Der vorliegende Beitrag untersuchte die Veränderung der ethischen Position durch die Teilnahme an einer Lehrveranstaltung zu dem Bereich IuG, in der ethische Aspekte – wie auch an anderen Hochschulen üblich neben Inhalten und Lernzielen aus gesellschaftlichen, ökonomischen, rechtlichen und politischen Kontexten – im Informatikstudium maßgeblich adressiert werden. Dabei musste festgestellt werden, dass sich durch die Teilnahme nur bedingt Veränderungen der ethischen Position feststellen ließen. Die Ergebnisse können somit als Plädoyer für eine explizitere Behandlung ethischer Aspekte in

Lehrveranstaltungen aus dem Bereich IuG verstanden werden. Außerdem existiert rund um die dargestellte Forschungsfrage erheblicher weiterer Forschungsbedarf. In folgenden Durchläufen soll dabei stärker inputorientiert vorgegangen werden und die Lehrveranstaltung um weitere Vorlesungsinhalte, u. a. zu ethischen Grundlagen, ergänzt werden. Wünschenswert wäre außerdem die Entwicklung von (kürzeren) Interventionen (wie bspw. Workshops, E-Learning-Szenarien, Serious Games, ...) zur Ausbildung einer ethischen Position bzgl. der Entwicklung von Informatiksystemen, die sich bspw. in der Weiterbildung oder auch im schulischen (Informatik-)Unterricht einsetzen ließen. Diese sollten ebenso wie zukünftige Lehrveranstaltungen wissenschaftlich begleitet werden; das Untersuchungsinstrument könnte dabei um weitere Konzepte, die bspw. auch Holman u. a. untersuchen, ergänzt werden, so bspw. um den Optimismus bzw. Pessimismus hinsichtlich des zukünftigen Einflusses von Informatiksystemen auf die Gesellschaft [Ho22]. Auch die Motivation und das Interesse an ethischen bzw. sozioinformatischen Aspekten im Allgemeinen sowie mit Bezug auf konkrete Fragestellungen und Aspekte sollen dabei stärker in den Blick genommen werden. Methodisch könnten dabei Mixed-Methods-Ansätze (also unter Einbezug der Erhebung und Auswertung qualitativer Daten) in Betracht gezogen werden. Eine weitere zu schließende Forschungslücke betrifft grundsätzlich kognitive Lernvoraussetzungen im Sinne von Studierendenvorstellungen zu ethischen Aspekten in der Informatik, die bislang (unserem Kenntnisstand nach) unerforscht sind.

4 Literaturverzeichnis

[BSM18] Brandenburg, S.; Schott, R.; Minge, M.: Zur Erfassung der ethischen Position in der Softwareentwicklung. In (Dachselt, R.; Weber, G., Hrsg.): Mensch und Computer 2018 - Workshopband. Gesellschaft für Informatik e.V., Bonn, 2018.

[Ca17] Capurro, R.: Digitization as an ethical challenge. AI & Society, 32(2):277–283, 2017.

[GSM21] Große-Bölting, G.; Scheppach, L.; Mühling, A.: The Place of Ethics in Computer Science Education. In (Desel, J.; Opel, S.; Siegeris, J., Hrsg.): 9. Fachtagung Hochschuldidaktik Informatik (HDI) 2021. FernUniversität in Hagen, Hagen, S. 85–91, 2021.

[Ho22] Horton, D.; McIlraith, S. A.; Wang, N.; Majedi, M.; McClure, E.; Wald, B.: Embedding Ethics in Computer Science Courses: Does It Work? In: Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 1. SIGCSE 2022, ACM, New York, S. 481–487, 2022.

[Hü21] Hübner, D.: Einführung in die philosophische Ethik. Vandenhoeck & Ruprecht, Göttingen, 3. Auflage, 2021.

[PPH13] Paech, B.; Poetzsch-Heffter, A.: Informatik und Gesellschaft: Ansätze zur Verbesserung einer schwierigen Beziehung. Informatik-Spektrum, 36(3):242–250, 2013.

[St21] Stavrakakis, I.; Gordon, D.; Tierney, B.; Becevel, A.; Murphy, E.; Dodig-Crnkovic, G.; Dobrin, R.; Schiaffonati, V.; Pereira, C.; Tikhonenko, S.; Gibson, J. P.; Maag, S.; Agresta, F.; Curley, A.; Collins, M.; O'Sullivan, D.: The teaching of computer ethics on computer science and related degree programmes. a European survey. International Journal of Ethics Education, 2021.

We are smart, smarter, the smartest?! Smart systems beyond optimization

Karolin Eva Kappler[1], Katharina Ebner[2], and Stefan Smolnik[3]

Abstract: Smart systems are not just technical tools but endowed with agency and discretion. Whereas their character as a black box is widely discussed, their moral inaccessibility is not. Broadening the issue of 'smartness', our research concern is as follows: How is the complex, typically ambiguous and contradictory moral infrastructure of late modern societies mirrored and eventually transformed by smart and learning systems? In order to conceptualize this question, we combine the normative reasoning of the Economics of Convention (EC), which distinguishes several competing moral orders, with the assumed smartness of learning systems. We propose to confront smartness with conflicting moral registers and to revise the competences which enable learning systems and humans to interact in a common "space of negotiation."

Keywords: Smartness, learning systems, Economics of Convention, competences

1 Introduction

The smartness of things is a topical issue and there is the widespread believe that smart systems will help to make things and systems smarter and begin to penetrate ever more spheres of society, from everyday social recommendations to life-changing decisions [BH18]. But technologies under the umbrella of 'smartness' are not just technical tools. Rather, they are growing into a new role as competent actors, equipped with learning capabilities, endowed with agency and discretion [Hu20]. Smart systems seen as "dark matter (figuratively speaking) which is holding societies together without being seen or being properly recognized by humans" [KV18] sustain the thesis of a fundamental mutual inaccessibility between smart systems and humans. This is consistent with AI/ML literature: Contemporary AI/ML systems are complex and difficult to understand, even for experts [Ba21]. Further, interactive learning frameworks such as active learning [Se12] and coactive learning [SJ15] focus on the optimization of systems and users and lack a sociological framework capable of dealing with the concept of competent actors and inconsistent or contradictory moral justifications so typical for late modern societies.

[1] FernUniversität in Hagen, Fakultät für Wirtschaftswissenschaft, Lehrstuhl für Betriebswirtschaftslehre, insb. Betriebliche Anwendungssysteme, 58084 Hagen, karolin.kappler@fernuni-hagen.de
[2] FernUniversität in Hagen, Fakultät für Wirtschaftswissenschaft, Lehrstuhl für Betriebswirtschaftslehre, insb. Betriebliche Anwendungssysteme, 58084 Hagen, katharina.ebner@fernuni-hagen.de
[3] FernUniversität in Hagen, Fakultät für Wirtschaftswissenschaft, Lehrstuhl für Betriebswirtschaftslehre, insb. Betriebliche Anwendungssysteme, 58084 Hagen, stefan.smolnik@fernuni-hagen.de

doi:10.18420/inf2022_58

In this sense, we are caught in the trap of thinking users are stupid, or at least incompetent, and systems are getting smarter. Hence, questions on the competence of users, but also systems and machines arise, as well as a debate on the 'right' smartness of things and humans [FPS18, WSH18].

Smartness in the context of smart systems refers to the use of information and communication technologies (ICT) and displays some level of intelligence. Enabled by algorithms, sensors and actuators, or interfaces for machine-to-machine (M2M) or human-to-machine (H2M) communication, smart systems – as we understand them in this paper – show an observable targeted or adaptive behavior [KWH13]. Another, more mathematical, characteristic of a smart system, which is of particular relevance here, is the notion of a goal function, also often referred to as win/loss function. Technically speaking, any smart system is seeking for an optimal solution state, which is often a local optimum displayed as one goal function. Since a smart system has only one such goal function seeking for one type of optimal state, from the perspective of pluralistic orders of worth, the decision capacity and thus the "true" competence of smart systems have to be challenged and criticized.

Consequently, smartness understood as optimization, reduces the plurality of orders of worth. Against this background, we propose a 'strong' program of normative reasoning to learning systems: the Economics of Convention (EC; see BT06). From this viewpoint, society is characterized by an incommensurable plurality of orders of worth or "registers". Each of these registers has its own normative criteria or "grammar" of critique and justification. Consequently, smartness is not a property of a social actor, let alone an algorithm, but has to be continually negotiated and mediated. Therefore, after a short overview over the state-of-art of smart systems, we propose a paradigmatic shift in understanding competence and smartness.

2 Smartness reloaded or the plurality of smartness(es)

As seen in the previous paragraphs, the common understanding of "smartness" is often reduced to the optimization of specific aspects supported by a system or application. In contrast, we propose a normative approach that has never been applied to the field of smart systems before: the EC (BT06). By introducing a pluralist approach, the EC presents a general theory to explain, initially, the organization of economic activity [SS97] and the generation of statistical categories [De98]. Boltanski and Thévenot [BT06] identified six conventions, also called registers, each of which is based on Western liberal societies' various philosophical foundations and concepts of worth and a common good: Industry, Market, Inspiration, Civic, House, and Opinion. This list was expanded by adding two more registers, Project and Green [BC13, LT00]. Table 1 provides an overview of each of these registers with their justification principles.

Register	Common good	Values
Industry	Increased efficiency	Functionality, optimization
Market	Economic growth	Competition, cost, profit
Inspiration	Inspiration	Emotion, spontaneity
House	Tradition	Hierarchy, trust
Civic	Collective will	Inclusivity, equality
Opinion	Public opinion	Popularity, fame
Project	Innovation and network	Activity, connection
Green	Protection of environment	Environmental activism

Table 1: Registers of worth

According to Table 1, there is a plurality of possible registers, that actors develop and manage in order to evaluate and coordinate situations [BT06]. Actors can, depending on their interactions with others, easily pass from one register to another. Each register acts as a logical, harmonious order of statements, objects, and people that provides a general sense of justice. The typology offers an applicable framework for identifying the plurality of registers, as well as conflicts with other registers. Further, the EC assumes that actors in situations are themselves competent to assess the appropriateness of conventions as normative logics for coordination, evaluation, and interpretation. According to this understanding, actors cannot be stupid, but their supposed rationality follows different registers, which might differ from the logics of efficiency and optimization (according to the Industry register) or profit maximization and cost reduction (according to the Market register). They might be guided by their pursuit of fame (according to the Opinion register) or their claim for solidarity and equal rights (according to the Civic register). Systems (and organizations) which do not consider this plurality of registers might label users as stupid, as, e.g., they do not follow the common good of increased efficiency claimed by the system. Hence, a mutual understanding of the (legitimate) plurality of registers and the competence of actors and things is necessary.

3 Competences and plurality revised: spaces of negotiation

We argue that to overcome the conflicting positions that humans pose towards smart systems, we need to re-consider the way humans' and machines' competences are regarded. We argue that we need a new conceptual – and possibly technical – realm that opens up room for humans and machines to "truly" interact. We refer to this room as a "space of negotiation". In the following, we will briefly illustrate our ideas how such a space of negotiation might manifest.

3.1 Plurality of orders of worth

From this perspective, smart systems and humans can be considered as equitable actors that both can and should learn from each other [Hu20, FPS18]. The major shift we suggest

here in relation to established understandings of traditional ML, in which learning refers to adding new data that will lead to a refinement of the decision model, is that we expect smart systems to also consider a plurality of orders of worth in their decision-making – just as human actors would do. This position stresses and reflects the challenge that many learning systems can at best come up with a local optimum – and that sometimes this local optimum might not be even applicable since it acts on screwed, biased, or even false assumptions.

In a first approach, Solans et al. [So21] have performed an automatic classification of registers to detect the dominant registers and moral orders in the AI-field. They have found, e.g., an ongoing conflict between the Industrial and House register around AI, the first referring to the concepts of standardization and optimization linked to AI, while the latter reflects the debates on the hierarchy between humans and AI and trustworthiness. The automatic detection of registers, as proposed by the classifiers, is able to shed light on the underlying moral assumptions in the AI (and other) fields. By this, it supports a deepened and mutual understanding of different points of view and moral backgrounds. As shown by this initial research, smart systems can be trained to detect and classify registers.

3.2 Spaces of negotiation

Consequently, the moral classifiers enable the democratization of smart systems by putting humans into the loop of automatic machine and system learning through an explanatory interactive learning approach [TK19], in which the users can see and correct a system's prediction and its explanation (based on the different moral registers) in a common space of negotiation. Through the mutual transparency regarding the underlying moral registers of learning systems' (algorithmic) decisions and their acceptance by humans, not only a deeper understanding can be obtained, but also possible conflicts can be made visible and hence negotiable. In this sense, the underlying target function becomes part of the space of negotiation, by differentiating the related optimization goal according to the cooperative and conflictive potential shown by the registers. We argue that by introducing the logic of spaces of negotiation, one of the main ethical problems of smart systems can be addressed: their paternalistic character, which incapacitates the users (at least partially), as the learning system takes decisions interfering with their lives [Hu20, De19]. A common space of negotiation might empower and democratize the process, by offering more transparency and enhancing the mutual understanding of the underlying moral orders of both – the smart system and the user(s).

Through this interdisciplinary approach and the integration of EC-grounded criteria, (required) competences change not only for the smart systems, which are confronted and supposed to learn to deal with the plurality of orders of worth and competent users. Humans will also have to adapt to interact with (partly) transparent and accountable systems. Analytically, new ways of connecting humans and smart systems by sensitizing the latter for registers and their moral fabric can be explored, leading to an entirely new

level of human-machine interaction within (new) spaces of negotiation.

4 Smartness revised: We are more than smart, we are different!

Smart systems profoundly shape individual choices and public judgement, on the one hand facilitating decisions and the coordination of actors, on the other obscuring their innards to most people, if not all of us. They therefore are increasingly taking part in distributing resources, attention and life chances, thereby gaining social power. So, what is at stake (from a technical and social point of view)? We assume three possible paths/ideal types for the co-evolution of smart systems and society: an *authoritarian* one, wherein choices, interests and prejudices are built into the devices, inaccessible to humans; second, a *rejection* of algorithmic power and ordering just because these devices are just that: unjust and arcane; thirdly, a *participatory* path wherein smart systems are open for negotiation and re-negotiation. The above mentioned paradigmatic shifts claim to reshape the interactions between humans and machines in a participatory way, by contributing to a broader understanding of the plural moral tissue of society and its underlying assumptions.

Hence, we propose to consider implementing different forms of "smartness" in systems and carry out pilots to test these new implemented taxonomies in running smart systems, e.g. smart cities, ranking systems etc. Further, we ask to discuss and reflect, which valuation orders and systems will be needed in a short/middle/long-term perspective, considering the proved complexity to define common goals for society or specific groups within society. Consequently, our approach pushes the issue of an estrangement of humans and systems further by addressing not only a smart systems's cognitive, but rather its moral inaccessibility.

Bibliography

[Ba21] Bauer, K. et al.: Expl(AI)n It to Me – Explainable AI and Information Systems Research. Business & Information Systems Engineering 2/63, pp. 79–82, 2021.

[BC13] Boltanski, L.; Chiapello, È.: Der neue Geist des Kapitalismus. UVK, Konstanz, 2013 [1999].

[BH18] Brougham, D.; Haar, J.: Smart Technology, Artificial Intelligence, Robotics, and Algorithms (STARA): Employees' perceptions of our future workplace. Journal of Management & Organization 2/24, pp. 239–257, 2018.

[BT06] Boltanski, L.; Thévenot, L.: On Justification. Economies of Worth. Princeton University Press, Princeton, Oxford, 2006.

[De19] Dellermann, D. et al.: Hybrid Intelligence. Business & Information Systems

Engineering 5/61, pp. 637–643, 2019.

[De98] Desrosières, A.: The Politics of Large Numbers. A History of Statistical Reasoning. Harvard University Press, Cambridge, London, 1998.

[FPS18] Faraj, S.; Pachidi, S.; Sayegh, K.: Working and organizing in the age of the learning algorithm. Information and Organization 1/28, pp. 62–70, 2018.

[Hu20] Huysman, M.: Information systems research on artificial intelligence and work: A commentary on "Robo-Apocalypse cancelled? Reframing the automation and future of work debate". Journal of Information Technology 4/35, pp. 307–309, 2020.

[KV18] Kappler, K.; Vormbusch, U.: 'Value-veillance': The Social Practices of Valuing Surveillance in Emergency Management. In (Boersma, K.; Fonio, C. Eds.): Big Data, Surveillance and Disasters. Routledge, London, New York, pp. 179–197, 2018.

[KWH13] Kagermann, H.; Wahlster, W.; Helbig, J.: Umsetzungsempfehlungen für das Zukunftsprojekt Industrie 4.0 – Abschlussbericht des Arbeitskreises Industrie 4.0, Berlin, 2013.

[LT00] Lamont, M.; Thévenot, L. Eds.: Rethinking comparative cultural sociology. Repertoires of evaluation in France and the United States. Cambridge University Press, Cambridge, 2000.

[Se12] Settles, B.: Active Learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 2012.

[SJ15] Shivaswamy, P.; Joachims, T.: Coactive Learning. Journal of Artificial Intelligence Research 53, pp. 1–40, 2015.

[So21] Solans, D. et al.: Learning to Classify Morals and Conventions: Artificial Intelligence in Terms of the Economics of Convention. Proceedings of the International AAAI Conference on Web and Social Media 1/15, pp. 691–702, 2021.

[SS97] Storper, M.; Salais, R.: Worlds of production. The action frameworks of the economy. Harvard University Press, Cambridge, 1997.

[TK19] Teso, S.; Kersting, K.: Explanatory Interactive Machine Learning. In (Conitzer, V. Ed.): Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery, New York,NY,United States, pp. 239–245, 2019.

[WSH18] Waardenburg, L.; Sergeeva, A.; Huysman, M.: Hotspots and Blind Spots. Springer, Cham, pp. 96–109, 2018.

Crisis escalation from autonomous weapon systems: agent-based modeling for military-technology assessment

Ole Lemke[1], Jürgen Altmann[2], Martin Degeling[3]

Abstract: Autonomous weapon systems (AWS) are not yet established, but strong military motives exist to develop and introduce them. A major fear is that they could deteriorate the stability in a crisis between two actors. To investigate this problem, we use agent-based modeling and simulation (ABMS). Here one AWS fleet patrols the perimeter of a protected area while a second moves outside. AWS of the first shoot if they sense being attacked or if AWS of the second enter the protected area. Errors in attack detection or in position determination can lead to escalation from crisis into war. The escalation probability is evaluated with a variation of these errors. Moreover, we elaborate on the shortcomings of ABMS in general.

Keywords: autonomous weapon systems, military-technology assessment, preventive arms control, agent-based modeling and simulation, crisis escalation

1 Introduction

In the last 20 years, uninhabited armed vehicles have spread rapidly. They attack under remote control by a human operator. But military research and development work for the next step, autonomous weapon systems (AWS). Without international limitations, the world could see also a similar proliferation in AWS. The introduction of AWS would affect international security. In particular, there are fears that AWS could have a negative impact on crisis stability. To better understand the consequences and develop preventive arms control, military technology assessment is necessary.

We approach the topic using agent-based modeling and simulation (ABMS). We (a) assess the probability of crisis escalation in a simple scenario and (b) reflect on whether ABMS is a helpful tool for investigating such a scenario.

2 Autonomous weapon systems and crisis stability

While present armed uninhabited air vehicles (UAVs) can perform some activities automatically, such as taking off, landing or reaching a target region, a human operator

[1] Ruhr-Universität Bochum, Alumni, ole.lemke@rub.de,
[2] Technische Universität Dortmund, Fakultät Physik, August-Schmidt-Str. 4, 44227 Dortmund, juergen.altmann@tu-dortmund.de
[3] Ruhr-Universität Bochum, Lehrstuhl für Systemsicherheit, Universitätsstrasse 150 ID 2/439, 44780 Bochum, martin.degeling@ruhr-uni-bochum.de

doi:10.18420/inf2022_59

remotely controls the specific task of attacking. An AWS, on the other hand, would attack without human control. According to the International Committee of the Red Cross [IC16] an AWS is:

> Any weapon system with autonomy in its critical functions. That is, a weapon system that can select (i.e. search for or detect, identify, track, select) and attack (i.e. use force against, neutralize, damage or destroy) targets without human intervention.

Remotely controlled, uninhabited armed vehicles already bring military benefits: they can save cost, be deployed in large numbers and be smaller. AWS promise additional advantages: no remote-control link can be jammed or betray the presence, and, above all, transmission delays and human reaction time are avoided. AWS pose serious problems, however. Can they comply with the laws of warfare, e.g., discriminate between combatants and civilians? How will they react in circumstances not foreseen in their programming? How will AWS interact with enemy AWS?

Strategic stability is decisive in preventing war between potential adversaries. It exists if neither side could gain significantly from attacking. One aspect is crisis stability: if even a fast pre-emptive attack could not prevent an unacceptable outcome, it will not be carried out by a rational actor.

There are several examples from the Cold War where the detection of enemy attack turned out to be a false alarm, , and only human and rational decisions averted launches of nuclear missiles as a second strike. A famous example is Lieutenant Colonel Stanislav Petrov in the Soviet Union. On Sept. 26, 1983, Petrov correctly classified a message about an intercontinental ballistic missiles (ICBM) attack as a false alarm, possibly preventing a nuclear strike that would have been the first one [Ho99]. Here the human in the loop acted as a fail-safe.

In AWS, the human element would be removed. In particular, in a crisis with two AWS fleets at a short distance, wrong indications of attack (sun glint perceived as a rocket launch, computer malfunction, misinterpretation of an adversary action) could lead to shooting and start an unintended “flash” war [AS17; Sc20].

3 The experiment

Agent-based modeling and simulation (ABMS) is a standard tool used for technology assessment. The ABMS tries to model sociotechnical processes in computer simulations. In these simulations, several agents interact with each other, including making autonomous decisions [WR17]. ABMS mainly consists of three components, the agents, the rule set, and the topology. Military studies have modeled AWS [ME1-21] but have not looked at stability.

A possible future scenario for modeling must meet some requirements. It should be relevant, possible to be modeled, and not too hypothetical. In the present case, the deployment scenario should be a relevant example of conflict stability. The individual weapon system's concrete effectiveness or technical superiority is less critical here. What matters is whether the deployment can lead to unintended escalation. The operational scenario for this short study must allow substantial reduction and not require too many hypothetical assumptions. Therefore, all weapon systems should exist in a similar form and be based on current and past deployments of UAVs.

Realistic operational scenarios can be found through literature research and a look at recent military conflicts with the increased use of UAVs. During the Nagorno-Karabakh war in 2020, UAVs (still under remote control) were used to attack battle tanks, artillery positions, or surface-to-air missile (SAM) systems [WK20]. In a future conflict, AWS could be used similarly. For example, AWS could observe an area and detect specific weapon systems with sensors, select targets and engage them. However, they could also be used to defend an area, objects, or troops against such attacks [AH19].

AnyLogic is a widespread simulation tool and is suitable for ABMS. It combines system dynamics (SD) modeling, discrete-event (DE) modeling, and agent-based (AB) modeling [Bo14]. The software is proprietary, but a “Free Personal Learning Edition” is available.



Figure 1: Snippet of the real-time view of the simulation. The AWS are shown strongly magnified.

The AWS were modeled on the General Atomics MQ-9 Reaper for the simulation. The model AWS are equipped with an air-to-air missile to engage enemy AWS. The cruise control system lets the AWS fly along waypoints in a geographic information system (GIS) with constant speed. After startup, the cruise control system is in the "patrol" state on a linear course. The course is corrected once per second. This loop is only terminated by reaching the target GIS point or by the destruction of the agent. During the flight, the AWS permanently searches for other nearby aircraft. The detected aircraft are classified. Aircraft classified as hostile are observed and evaluated. An attack occurs when a hostile aircraft is in a predefined GIS area or launches an attack.

With the cruise control system and the search-select-attack loop, the combat OODA loop (Observe Orient Decide Act) is completely defined. The AWS can search for and detect targets, decide whether to engage them and then engage them. The air-to-air missiles are

also modeled as agents. They function as fire-and-forget weapons. Once launched, they fly to their target. They make a course correction every 50 ms. If the missiles do not reach their target, they destroy themselves.

In the experiment, a larger and a smaller fleet of AWS face each other (figure 1). The larger AWS fleet forms a defensive chain at the edge of the territory. The smaller AWS fleet flies past this area in international air space, but does not penetrate the territory. There are two conditions for an escalation; one is when an AWS detects firing by an enemy AWS and the other is when an AWS detects an enemy AWS approaching the

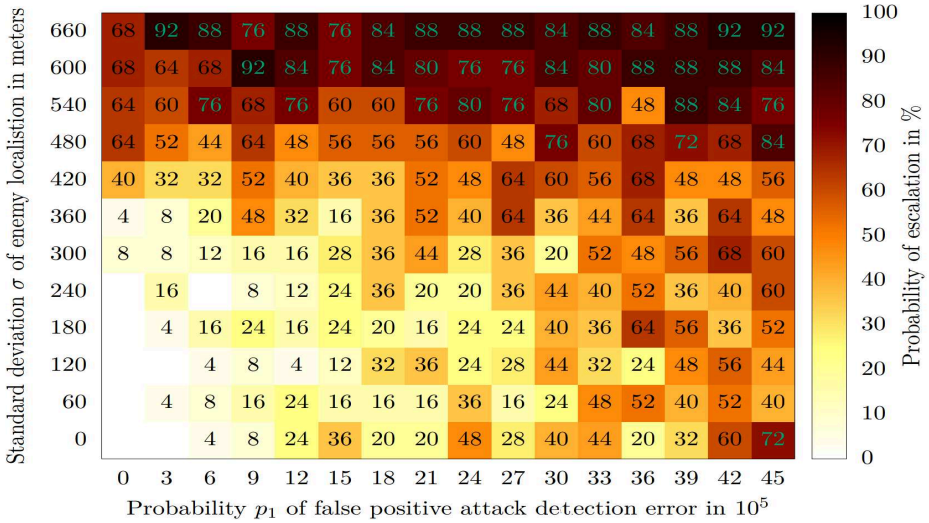


Figure 2: Probability of escalation as a function of the probability of false positive attack detection error and the standard deviation of enemy localization in meters.

sovereign area to less than 100 m.

An unintentional escalation should not be triggered without errors in the sensors or in the behavior of the AWS. However, if the attack detection and the position determination of the enemy AWS are faulty, escalation can occur. The errors used in this work are located in the OODA loop. Faulty sensor data lead to a wrong decision. Two error sources were implemented, both in the select function of the search-select-attack loop.

The first concerns attack detection. The system returns false positives with a probability p_1. This can cause an enemy agent to be incorrectly classified as attacking and countermeasures, here firing, can be initiated. The second error concerns the positioning of hostile - agents. Determining the distance to the airspace to be protected is subject to a normally distributed error with a standard deviation σ. An enemy agent can be incorrectly classified as entering or having entered the airspace to be protected and countermeasures, again firing, are initiated. One simulation using random numbers for

both parameters takes 1h 45m in simulation time with, on average, 33 patrolling aircraft. The experiment is executed 25 times in each iteration with identical error settings.

The result of the ABMS gives the probability of escalation in dependence on the probability of false-positive attack detection error and the standard deviation of enemy localization (figure 2). With no error on both, the probability of escalation will be 0. With increasing error in one of both systems, the probability of escalation rises. In parallel, the escalation probability was calculated analytically using the same probability and error distribution. Here the 2-D result function is smoother. The standard deviation of the difference between both distributions is 4%, measured in the units of the probability of escalation.

4 Discussion

Thesimulation evaluation shows that even minimal errors in the attack detection system, on a scale of 1:10,000, can lead to an unintended escalation in somewhat realistic mission scenarios. The localization error increases the probability. The present model has limitations. The interacting behavior of different AWS types could not be investigated. Moreover, it was impossible to identify how different errors in sensor technology or in error handling interact with each other and reinforce each other. This is because the simulation implementation becomes more and more complex, and the assumptions about the AWS implementation become more and more hypothetical. There is a general problem in finding a suitable model for military conflicts. The environment, the timing, the equipment of the opponents, and in particular, their actions are highly variable and hard to predict. The parties involved in the conflict have incentives to add to this complexity to surprise the adversary [Da18].

In the present model, ABMS have limited applicability in terms of behavioral analysis. Analytical computations produce similar results with much less effort. The question remains whether more complex modeling incorporating more accurate and realistic deployment can increase meaningfulness. However, every modeling, whether analytical or simulated, faces the same problem. The more complex scenarios become, the more difficult it is to find a simple and meaningful model which can predict the “qualitative” behavior of reality. Without verification of the results by measurement, a theory is not conclusive. However, experimentation with armed conflict is virtually impossible.

On the other hand, modeling and simulation can give some indications about the relationships between AWS and crisis stability in a simple form and make them accessible. Moreover, modeling and simulation seem to be the only possible “experimental” way to test theories about the behavior of AWS in conflict situations. However, there is a risk that the contradiction between the imperfection of the underlying model and the complexity of reality will not be seen by policymakers.

These problems also apply to other areas of military-technology assessment. We think that more precise models will be useful to better judge on the utility of ABMS, to look for indications of unintended behavior, and above all, to exclude it as far as possible.

5 Conclusions and outlook

Following literature evidence on the possible negative impact of autonomous weapon systems (AWS) on crisis stability, agent-based modeling and simulation (ABMS) of a hypothetical deployment of AWS was created and results were verified by analytical calculations. The simulation evaluation showed that even minimal errors in the sensor system could lead to an unintended escalation in realistic mission scenarios. The simulation ended with the first shot. The further course of events was not modeled and therefore could not be analyzed. ABMS, in terms of behavioral analysis, appears to have limited applicability in a simple scenario. However, ABMS is one of the few ways to test theories about the behavior of AWS. However, there is a risk that the contradiction between the imperfection of the model and the complexity of reality is not sufficiently taken into account. More in-depth analyses seem required to investigate AWS's impact on crisis stability further.

Bibliography

[AH19] Amt für Heeresentwicklung. Künstliche Intelligenz in den Landstreitkräften - Ein Positionspapier des Amts für Heeresentwicklung, 2019.

[AS17] Jürgen Altmann and Frank Sauer. Autonomous weapon systems and strategic stability. Survival, 59(5):117–142, 2017.

[Bo14] Andrei Borshchev. Multi-method modelling: AnyLogic, John Wiley & Sons Ltd, 2014.

[Da18] Richard Danzig. Technology Roulette - Managing Loss of Control as Many Militaries Pursue Technological Superiority. Center for a New American Security, 2018.

[Ho99] David Hoffman. 'I Had A Funny Feeling in My Gut'. Washington Post, 1999.

[IC16] International Committee of the Red Cross (ICRC). Views of the International Committee of the Red Cross (ICRC) on autonomous weapon system, 2016.

[ME14-21] International Conference on Modelling and Simulation for Autonomous Systems (MESAS), 2014-2021.

[Sc20] Paul Scharre. Autonomous Weapons and Stability. PhD thesis, King’s College London, 2020.

[WK20] Jack Watling and Sidharth Kaushal. The Democratisation of Precision Strike in the Nagorno-Karabakh Conflict, 2020.

[WR17] Johannes Weyer and Michael Roos. Agentenbasierte Modellierung und Simulation: Instrument prospektiver Technikfolgenabschätzung. TATuP, 26(3):11–16, 2017.

Informatisches Gestalten unter Bedingungen von Gegenwartskomplexität und Zukunftsunsicherheit

Plädoyer für eine Verständigung in und mit Gesellschaft

Sercan Sever[1]

Abstract: Akteur:innen, die sich an der Gestaltung von Technik beteiligen, sind einer irreduziblen soziotechnischen Komplexität und Zukunftsunsicherheit ausgesetzt. Die Informatik als Gestaltungswissenschaft unterliegt in ihrem Handeln denselben Bedingungen, ist darüber hinaus tiefgreifend sozial wirksam, spätestens wenn ihre Entwicklungsresultate in die Gesellschaft diffundieren. Sie kann mit dieser Einsicht und ihrer Verortung in Gesellschaft diese als Chance begreifen und sich mit Gesellschaft verständigend gegenwärtige Orientierungsprobleme bewältigen. Dafür muss sich von unidirektionalen Vorstellungen von Wissenschaftskommunikation einerseits, von technokratischen Visionen andererseits verabschiedet werden. Durch Einsicht, Lossagung und Verständigung kann ein transformatives Potenzial der Informatik zur Bewältigung großer Herausforderungen für die Zukunft erwachsen.

Keywords: Gestalten, Verständigung, soziotechnische Komplexität, Zukunft, Unsicherheit

1 Einleitung

Die Informatik ist, zumindest in Teilen der Disziplin, ihrem reflexiven Selbstbild nach Gestaltungswissenschaft, die inzwischen tiefgreifend und weitreichend Bereiche der Gesellschaft mittels Entwicklung von Informationssystemen zu gestalten in der Lage ist wie bereits von Wolfgang Coy festgestellt wurde [Co92]. Mit einem wachsenden öffentlichen Bewusstsein für Digitalisierung, damit emergierenden Professionen und Qualifikationen aus dem Bereich der Informatik und Zukunftsszenarien einer mehr und mehr digitalisierten Gesellschaft scheint ihr diese gestaltende Verfasstheit auch gesellschaftlich zugesprochen zu werden. Gleichzeitig sehen wir uns heute vor eine soziotechnische Komplexität gestellt, die nicht nur alltägliche Techniknutzer:innen nicht, sondern kaum ein:e Informatiker:in oder Ingenieur:in als Technikgestalter:innen im engeren Sinne umfänglich verstehen können [Gr15]. Hinzukommt, dass die Zukunft mit heutigen technischen Mitteln – vielleicht mehr denn je – als ein offener Raum von Möglichem hervortritt, den zu prognostizieren beinahe unmöglich ist, zu bestimmen mindestens mit Unsicherheitszuwächsen in Kauf genommen werden muss. Es stellt sich daher die Frage, wie sich informatisches Gestalten unter Bedingungen von

[1] Martin-Luther-Universität Halle-Wittenberg, ZSB, Franckeplatz 1, Haus 31, 06099 Halle (Saale), sercan.sever@zsb.uni-halle.de

doi:10.18420/inf2022_60

Gegenwartskomplexität und Zukunftsunsicherheit überhaupt vollziehen lassen kann und soll. Dieser Beitrag will zur Beantwortung zunächst eine Gegenwartsbeschreibung soziotechnischer Komplexität entfalten, sodann epistemische Überlegungen zur Zukunft und des unsicheren Wissens darüber anstellen und für einen sich im Umgang daraus ergebenden Einbezug von Gesellschaft in informatische Gestaltungsfragen plädieren.

2 Soziotechnische Komplexität und Zukunftsunsicherheit

Die Unterscheidung von Gesellschaft und Technik ist angesichts heutiger Vernetzung und Verkopplung humaner und nicht-humaner Einheiten nicht mehr problemlos haltbar. Vielmehr dient sie der analytischen Betonung von sozialer, vermeintlich genuin menschlicher, und nicht-sozialer, vermeintlich genuin technischer, Aspekte, kann damit allerdings nicht die sich herausgebildete soziotechnische Totalität [Ro09] selbst in Gänze erfassen, denn bekanntlich ist das Ganze – die soziotechnische Realität – mehr als die Summe seiner Teile – Gesellschaft und Technik. Bereits auf dieser deskriptiven Ebene und einer damit einhergehenden Begriffswahl scheiden sich die Geister. Während lebensweltlich die Unterscheidung scheinbar einfach funktioniert, wir beispielsweise als Nutzer:innen einen instrumentellen Technikbegriff an konkrete Techniken anlegen können, ist sie in Disziplinen, die sich u.a. mit Technikgenese, -philosophie, -soziologie befassen, schon seit Längerem nicht mehr tragfähig [Ra99]. Nicht, weil in der Zwischenzeit geglaubt wurde das Verhältnis habe sich umgekehrt und Technik mit Agens ausgestattet könne nun das Soziale determinieren, sondern weil erkannt wurde, dass bei solchen einseitigen Zuschreibungen und voneinander losgelösten Betrachtungen der Einheiten etwas vom Ganzen verlorengehe, das die zu beschreibenden Phänomene jedoch wesentlich kennzeichne [Hö17]. Mittlerweile gibt es verschiedene Ansätze aus unterschiedlichen Disziplinen mit dem Zweck angemessener Beschreibung von Phänomenen des sozial-wie-technisch Ganzen. In diesem Beitrag wird die *soziotechnische* Begrifflichkeit bevorzugt [Ro09], weil in ihr die zwei Größen – Gesellschaft/Mensch und Technik – zusammenfallen und so ihre losgelöste Betrachtung bereits orthographisch verhindern. Entscheidender als die Festlegung auf eine bestimmte Begrifflichkeit ist aber, dass sich die soziale wie technische Komplexität der Gegenwart mittels der Überwindung binärer Formeln vor allem eingestanden wird, möchte auf diese hin etwas erkannt oder entwickelt werden. Es braucht ein „zeitgemäß komplexes Technikverständnis, das vor allem soziotechnische Komplexität verstehen hilft“, sprich das irreduzible und nicht simplifizierbare [Gr15] „Geflecht von Interdependenzbeziehungen, die einerseits gesellschaftliche Praxen und andererseits materielle und soziale Techniken umfassen“ [GJ05]. Verkennt man die Soziodimension des Technischen oder jene Technikdimension des Sozialen, wird eine unmögliche Komplexitätsreduktion möglich, wodurch nur eine verkürzte Sicht, ein lückenhaftes Wissen und letztlich nur verkürzte Deutungs- und Entwicklungsangebote begünstigt werden können. Diese Einsicht ist nicht nur theoretischer Art, sondern wirkt sich auf die Gegenstandskonstruktion, die Perspektive und die Reichweite gestalterischen Handelns aus, denn die soziotechnische Komplexität der Gegenwart hat mittelbar – über Handlungen in dieser – Folgen für die Zukunft. Wenn

die soziotechnische Komplexität der Gegenwart anerkannt wird, folgt zugleich die Einsicht der Undurchschaubarkeit der fortschreitenden Verkopplung soziotechnischer Phänomene: „Die soziale Komplexität wird durch immer höhere technische Komplexität ergänzt und beide potenzieren sich in soziotechnischer Verschränkung“ [Gr15]. Welcher Handlungsentwurf durch seine intendierte Realisierung letztlich zu welchem Handlungsergebnis führt, ist bei dieser komplexen Sachlage nicht einzuschätzen, geschweige denn unter Rekurs auf bezweifelbare oder zumindest in Auflösung begriffene Gesetzmäßigkeiten und nicht nachzuvollziehende Kausalitäten zu rekonstruieren, gar zu prognostizieren. Und doch muss und will entschieden werden, gerade unter Bezugnahme auf Zukunft. Sowohl handlungstheoretisch als auch einer modernen Begründungsstruktur nach möchten wir unsere Entscheidungen mit Gebrauch der Zukunft und in Anlehnung an Zukunft fällen. Handlungsentwürfe reduzieren künstlich die Offenheit der Zukunft soweit, bis auf Grundlage dieser so entstehenden „fiktionalen Erwartung“ entschieden werden kann, „*als ob* sich die Zukunft in der angenommenen Weise entfalten würde“ [Be15, Herv. i. O.]. Durch diesen Reduktionsprozess werden handlungstheoretisch Entscheidungen ermöglicht, auch solche im Rahmen eines planenden Gestaltungs- und Entwicklungshandelns. Wenn sich allerdings Handeln lediglich unter Bezugnahme auf künstlich reduzierte und damit gewiss gemachte Zukunft (Zukunfts“wissen“) vollzieht, dann ist „grundsätzlich die Unsicherheit des Handelns anzuerkennen [Gr09b]. Damit gehört „[d]as Auftreten von nicht-intendierten […] Nebenfolgen in der Folge intentionaler Handlungsvollzüge […] ganz allgemein zu jedem menschlichen Handeln“ [GJ05]. Dennoch werden durch Inanspruchnahme der im Grunde unsicheren Zukunft zusätzlich „Begründungen für Entscheidungen unter Bedingungen von Ungewissheit geschaffen […], damit Entscheidungen nicht beliebig erscheinen“ [Be15]: „Ausgehend von gegenwärtigen Orientierungsproblemen wird auf dem Umweg über Zukunftsdebatten Orientierung für heute gesucht“ [Gr09a]. Bei Feststellung der soziotechnischen Komplexität der Gegenwart und des ungewissen und daher verunsichernden Status von Zukunft zeigt sich eine verblüffende Spannung in Entscheidungssituationen: Es muss eine Zukunft oder ein Bündel von Zukünften fiktional festgelegt werden, wo keine feststehen kann. Wir sollten eigentlich nicht wissen können, was wir tun sollen [GH05]. Wo also alles für die Zunahme von Kontingenz einer zu treffenden Entscheidung und ihrer Auswirkung auf Zukunft spricht und daher deutlichere Orientierungsverluste zu erkennen sein müssten, kommen Entscheidungen angesichts einer solchen Situation dennoch nicht zum Erliegen.

3 Bewältigung durch Verständigung

Durch Bezugnahme auf Zukunft für gegenwärtig anliegende Entscheidungen wird die Spannung der Entscheidungssituation nicht aufgelöst, als vielmehr mit dem Orientierungsproblem – sich an *der* Zukunft orientieren müssen und wollen ohne es zu können – umgegangen. Gerade bei der informatischen Technikgestaltung, die von denselben komplexitäts- und kontingenzsteigernden Bedingungen in Entscheidungssituationen ausgehen muss, dürfte die Frage nach der Begründung schwerer

ins Gewicht fallen, fließen ihre Entscheidungsresultate schließlich ganz wesentlich in die Gesellschaft hinein. Es macht einen Unterschied, ob ein nicht-unternehmerisch motiviertes Individuum zum Hobby an technischen Entwicklungen bastelt oder aber ein professionelles Kollektiv in unternehmerischer Absicht Technikgestaltung betreibt, deren Resultate es gilt marktreif zu machen und in die Gesellschaft diffundieren zu lassen, damit sie sich etablieren und rentieren. Letztere haben, ganz im Sinne der Informatik als Gestaltungswissenschaft, erheblichen Einfluss auf gesellschaftliche Prozesse und insgesamt auf soziotechnische „Figurationen des Individuums wie auch der Gesellschaft“ [AAR17], auch wenn das Ausmaß dieses Gestaltungsprozesses selbst bei gewissenhaftester Reflexion nicht gänzlich einzuschätzen und angesichts der gegenwärtigen Komplexität und ungewissen Zukunft zu bestimmen ist. Wenn für informatisches Gestalten dieselben Bedingungen angenommen werden dürfen wie für nicht-informatisches, zudem informatisches Gestalten durch zur Diffusion angelegte Entwicklungsresultate als sozial wirksam gelten muss und soll, ohne jene in der Zukunft liegende soziale Wirksamkeit tatsächlich vorwegnehmen zu können, dann braucht es eine zunehmende Öffnung und Anpassung der (Sozio-)Informatik im Sinne einer Theorie der Informatik [Co92]. Informatisches Gestalten als Teil gesellschaftlicher Technikentwicklung kann „nicht länger als die Domäne einzelner Erfinder und Unternehmer verstanden werden, sondern gewinnt zunehmend den Charakter einer öffentlichen Angelegenheit“ [Ro09], sodass Gesellschaft nicht nur als Objekt stärker in informatische Forschungs- und Anwendungsfragen, sondern selbst als Ort der Verständigung über informatische Fragestellungen, Reichweite der Gestaltung und Technisierung in den Vordergrund rückt. Dem gesellschaftlichen Diskurs, wie auch immer dieser definiert und womit auch immer repräsentiert wird, kommt in Sachen informatischer Gestaltung so eine höhere Bedeutung zu als womöglich bisher [LRW18]. Während in sozioinformatischen Ansätzen gesellschaftliche Aspekte Einbindung erfahren, braucht es weiterhin eine inter- wie transdisziplinäre Selbstreflexion, um Gesellschaft nicht nur indirekt und vage in einer passiv-reagierenden Rolle der Kommentatorin, sondern unter bidirektionaler Perspektive als proaktiv-agierende Gestalterin einzubinden. Unter dieser Perspektive stellen sich weniger Fragen nach unidirektionaler Dissemination, wissenschaftskommunikativ ansprechender Darstellung von informatischen Forschungsergebnissen, Meinungsabfrage zu Entwicklungsresultaten oder nachträglicher Akzeptanzverschaffung für technische Neuheiten, als vielmehr nach Einbindung in den Forschungs- und Entwicklungsprozess selbst als Beteiligte auf Augenhöhe. Damit werden immer wieder neu befeuerte, nicht abschließbare Diskussionen um die Rolle, Funktion, Infrastruktur und den Zweck von Forschung und Entwicklung und von Repräsentation „der“ Gesellschaft nach sich gezogen. Dies auszutarieren bleibt in Anbetracht etablierter Forschungspraktiken eine schwierige, jedoch nicht minder nötige Aufgabe für Forscher:innen. Bei gegenwärtigen und zukünftigen Herausforderungen, die stets soziotechnischer Natur sind und nur von großen Unsicherheiten aller beteiligten Laien wie Expert:innen – eine Dichotomie, die nicht problemlos haltbar ist [HHN21] – begleitet werden können, führt, nicht zuletzt unter demokratisierenden Vorzeichen, kein Weg an einer grenzüberschreitenden Verständigung vorbei. Die Informatik als Gestaltungswissenschaft kann dies als Chance begreifen und hier als

verantwortungsbewusste, transformative Wissenschaft in Erscheinung treten [GHS21]. Statt Vereinfachungsversuchen von soziotechnischen Zusammenhängen und Vergewisserungsversuchen von Zukunft kann durch in der Verständigung gemeinsam herzustellendes „socially robust knowledge“ [No03] den negativen Folgen von Unsicherheitszuwächsen entgegengewirkt werden, damit das Orientierungsproblem bewältigbar wird. Die wiederholte Feststellung soziotechnischer Komplexität und Zukunftsunsicherheit erinnert uns daran und animiert zu einer Verständigung informatischen Gestaltens *in* und *mit* Gesellschaft, in deren Sphären die Informatik letztlich selbst verortet ist, vielleicht durch die stete Anwesenheit digitaler Artefakte noch nie so spürbar wie heute. Eine Ablehnung dieser Verständigung mag in szientistischen und technokratischen Visionen Wiederhall finden, zeugt allerdings von einem unzureichenden Verständnis der Komplexität, des nur fiktionalen Zukunftswissens, des epistemologisch gleichen Ausganspunkts von Wissenschaftler:innen und Nicht-Wissenschaftler:innen, der gesellschaftlichen Verortung und somit ohnehin immer vergesellschaftlichten Wissenschaft selbst [Gr18a]. Im Zweifelsfall gilt es daher noch „die Unsicherheit aller so zu vergrößern, so daß man sich nur noch verständigen kann“ [Lu92], denn „[e]ine auf Verständigung abzielende Kommunikation muß […] zunächst einmal Unsicherheit vermehren und das gemeinsame Wissen des Nichtwissens pflegen“ [Lu92]. Gelingen kann die Einsicht zur Verständigung nicht nur durch eine Verunsicherung des (Zukunfts-)Wissens, sondern ferner durch eine Umwertung der Unsicherheit selbst, aus der schließlich erst Gestaltungspotenziale herauswachsen: „Entscheiden kann man nur, wenn und soweit nicht feststeht, was geschehen wird“ [Lu92]. Die „Unsicherheit des Zukunftswissens […] verdankt sich“ nicht nur „grundlegenden Strukturen von Mensch und Gesellschaft und ist damit unüberwindlich“, vielmehr ist sie positiv gewendet zugleich auch „Ausdruck der Gestaltbarkeit der Zukunft“ [Gr18b]. In Fragen der Technikentwicklung kann eine gestaltungswissenschaftliche Informatik selbstbewusster inter- und transdisziplinäre Einladungen aussprechen und sich transparent, informierend, moderierend und gestaltend [GHS21] in und mit Gesellschaft verständigen. Der anzunehmende Vertrauensvorschuss im digitalen Zeitalter für Informatiker:innen scheint dafür eine geeignete Ausgangslage zu sein.

4 Literaturverzeichnis

[AAR17] Allert, H.; Asmussen, M.; Richter, C.: Digitalität und Selbst. Einleitung. In (Allert, H.; Asmusen, M.; Richter, C., Hrsg.): Digitalität und Selbst. Interdisziplinäre Perspektiven auf Subjektivierungs- und Bildungsprozesse, transcript, Bielefeld, S. 9-23, 2017.

[Be15] Beckert, J.: Imaginierte Zukunft. Wie fiktionale Erwartungen wirtschaftliche Dynamik vorantreiben. In (Max-Planck-Institut für Gesellschaftsforschung, Hrsg.): Jahrbuch. Forschungsbericht 2015, https://www.mpg.de/10420061/mpifg_jb_2015, S. 29-36, 2015, Stand 04.05.2022.

[Co92] Coy, W.: Für eine Theorie der Informatik! In (Coy, W.; Nake, F.; Pflüger, J.; Rolf, A.; Seetzen, J.; Siefkes, D.; Stransfeld, R., Hrsg): Sichtweisen der Informatik. Vieweg, Braunschweig u.a., S. 17-32, 1992.

[GH05] Gamm, G.; Hetzel, A.: Eine zeitgemäss-unzeitgemässe Philosophie der Technik. In (Gamm, G.; Hetzel, A., Hrsg.): Unbestimmtheitssignaturen der Technik. Eine neue Deutung der technisierten Welt, transcript, Bielefeld, S. 9-15, 2005.

[GHS21] Goller, A.; Holst, J.; Singer-Brodowski, M.: Transformative Wissenschaft. In (Philipp, T.; Schmohl, T., Hrsg.): Handbuch Transdisziplinäre Didaktik. transcript, Bielefeld, S. 347-356, 2021.

[Gr15] Gransche, B.: Vorausschauendes Denken. Philosophie und Zukunftsforschung jenseits von Statistik und Kalkül, transcript, Bielefeld, 2015.

[Gr09a] Grunwald, A.: Wovon ist die Zukunftsforschung eine Wissenschaft? In (Popp, R.; Schüll, E., Hrsg.): Zukunftsforschung und Zukunftsgestaltung. Beiträge aus Wissenschaft und Praxis, Springer, Berlin u.a., S. 25-35, 2009.

[Gr09b] Grunwald, A.: Zum Handlungsbegriff in Technikphilosophie und Technikethik. In (widerstreit-sachunterricht, Hrsg.): Handeln. https://www.itas.kit.edu/pub/v/2009/grun09a.pdf, S. 1-7, 2009, Stand 04.05.2022.

[Gr18a] Grunwald, A: Denken in Alternativen. In (Weiterbildung, Hrsg.): Weiterbildung und Politikberatung. Ausgabe 5, S. 24-26, 2018.

[Gr18b] Grunwald, A.: Lob der Unsicherheit. Plädoyer für offene Technikzukünfte. In (Langenfeld, M.; Redlich, T.; Weidner, R., Hrsg.): Unsicherheiten der Technikentwicklung. Cuvillier, Göttingen, S. 44-62, 2018.

[GJ05] Grunwald, A.; Julliard, Y.: Technik als Reflexionsbegriff. Überlegungen zur semantischen Struktur des Redens über Technik. In (Philosophia Naturalis, Hrsg.): Volume 42, Issue 1, S. 127-157, 2005.

[HHN21] Hoffmeister, T.; Hundt, M.; Naths, S.: Laien, Wissen, Sprache. Theoretische, methodische und domänenspezifische Perspektiven, de Gruyter, Berlin u.a., 2021.

[Hö17] Hörning, K.: Wissen in digitalen Zeiten. In (Allert, H.; Asmussen, M.; Richter, C., Hrsg): Digitalität und Selbst. Interdisziplinäre Perspektiven auf Subjektivierungs- und Bildungsprozesse, transcript, Bielefeld, S. 69-85, 2017.

[LRW18] Langenfeld, M.; Redlich, T.; Weidner, R.: Unsicherheiten der Technikentwicklung als gesellschaftliches und politisches Handlungsfeld. In (Langenfeld, M.; Redlich, T.; Weidner, R., Hrsg.): Unsicherheiten der Technikentwicklung. Cuvillier, Göttingen, S. 26-42, 2018.

[Lu92] Luhmann, N.: Beobachtungen der Moderne. Westdeutscher Verlag, Opladen, 1992.

[No03] Nowotny, H.: Democratising expertise and socially robust knowledge. In (Science and Public Policy, Hrsg.): Volume 30, Issue 3, S. 151-156, 2003.

[Ra99] Rammert, W.: Technik. Stichwort für eine Enzyklopädie, https://www.ts.tu-berlin.de/fileadmin/fg226/TUTS/TUTS_WP_1_1999.pdf, 1999, Stand 04.05.2022.

[Ro09] Ropohl, G.: Allgemeine Technologie. 3. Auflage, Vieweg, Braunschweig u.a., 2009.

Die ethischen Leitlinien der GI als Rahmen für Informatik und Gesellschaft in der Bachelorausbildung

Gabriele Kunau [1], Andrea Kienle [2]

Abstract: Dieser Beitrag präsentiert ein Lehrkonzept für die Bachelorausbildung im Fach Informatik und Gesellschaft (I&G) entlang der ethischen Leitlinien der GI, das nun seit 7 Semestern am Fachbereich Informatik der Fachhochschule Dortmund im Einsatz ist. Es integriert aktivierende Lehrformate und aktuelle Themen, die Studierende auch emotional ansprechen. Mit diesem Konzept vermitteln wir Inhalte, die anschlussfähig an andere Veranstaltungen sind und es den Studierenden erlauben, ein eigenes Ethos zu Fragen der Technologiegestaltung zu entwickeln. Die qualitative Evaluation zeigt eine gute Akzeptanz der Studierenden für das Format und die Inhalte der Veranstaltung. Zudem demonstrieren die Beiträge der Studierenden, dass die Lernziele erreicht werden.

Keywords: Ethische Leitlinien, Informatikausbildung, Lehrkonzept I&G

1 Einleitung

Seit beinahe 30 Jahren beschäftigt sich die I&G-Community mit ihrem Selbstverständnis. In dem ersten Lehrbuch für I&G, das 1995 erschienen ist, stellt Arno Rolf das Fach Informatik und Gesellschaft als gleichberechtigte fünfte Disziplin neben die Disziplinen theoretische Informatik, praktische Informatik, technische Informatik und angewandte Informatik: „Informatik und Gesellschaft – dieses Fachgebiet analysiert die Wirkungen des Einsatzes der Informatik in unterschiedlichen Bereichen und entwickelt Kriterien und Methoden zur Gestaltung sozialverträglicher Informatiksysteme.“ [Ro95, S. 6]. Auch wenn sich diese Einordnung nicht durchgesetzt hat, hat die inhaltliche Beschreibung jedoch in großen Teilen nach wie vor Bestand.

Als oftmals unbenotetes Fach, das kein schnell verwendbares Verfügungswissen vermittelt, muss sich I&G besonders um Akzeptanz bei den Studierenden bemühen. Dieser Beitrag präsentiert ausgehend von einem kurzen Rückblick (Abschnitt 2) und der Beschreibung der ethischen Leitlinien (Abschnitt 3) unser Lehrkonzept für I&G (Abschnitt 4) und Erfahrungen damit (Abschnitt 5). In Abschnitt 6 stellen wir offene Fragen zur Diskussion.

[1] Fachhochschule Dortmund, Fachbereich Informatik, Emil-Figge Str. 42, 44227 Dortmund, gabriele.kunau@fh-dortmund.de
[2] Fachhochschule Dortmund, Fachbereich Informatik, Emil-Figge Str. 42, 44227 Dortmund, andrea.kienle@fh-dortmund.de, , https://orcid.org/ 0000-0002-3236-3418

doi:10.18420/inf2022_61

2 I&G im Wandel der Zeit

Im Anschluss an die oben erwähnte Definition von Arno Rolf haben sich in den Folgejahren einige weitere Überlegungen zum Selbstverständnis und der daraus resultierenden inhaltlichen und organisatorischen Gestaltung der I&G-Lehre ergeben. Unsere eigene Definition orientiert sich an diesen Entwicklungen und stellt den Begriff sozio-technischer Systeme in den Mittelpunkt: „Das Fach Informatik und Gesellschaft hat sozio-technische Systeme zum Gegenstand: (1) Es entwickelt Theorien und Modelle zu ihrer Beschreibung. (2) Es analysiert und beschreibt die Wechselwirkungen zwischen der Informationstechnik und den sie nutzenden sozialen Systemen. (3) Es entwickelt Methoden zu ihrer Gestaltung, die neben der Technikgestaltung auch die Veränderung im sozialen System im Blick haben.“ [KK14, S. 36].

Das Selbstverständnis des Faches hat auch Implikationen auf die Gestaltung der I&G-Lehre. Bereits in dem ersten Lehrbuch aus 1995 fordern Domeyer et al. Leitlinien für eine Didaktik von I&G, die auf die Verschränkung von Theorie und praktischen Fällen und die Förderung der Sozialkompetenzen der Studierenden abzielen [DGSW95]. Methodisch ist eine stärkere aktive Einbindung der Studierenden gefordert. In unserer eigenen Historie beobachteten wir wenig später „Schwierigkeiten, den Zugewinn von Diskussions- und Argumentationsvermögen zum Gegenstand üblicher Leistungsüberprüfungen zu machen. [...]. Es sind geeignete Aufgabenstellungen zu finden, die argumentativ strukturiertes Reflektieren fordern und fördern.“ [HHM98, S. 90]. Ein Erfahrungsbericht von Engbring und Selde zur I&G-Lehre aus dem Jahre 2012 kommt zu dem Schluss, dass I&G eine Pflichtveranstaltung im Bachelorstudium sein sollte und unter anderem auch die Historie der Informatik in den Blick nehmen sollte [ES12].

Eben diesen Ansatz verfolgen wir an der Fachhochschule Dortmund. Zur Unterstützung unserer Pflichtveranstaltung thematisieren wir in unserem Lehrbuch methodische Grundlagen zur Beschreibung und Gestaltung sozio-technischer Systeme in 10 Einheiten [KK14]. In der Zeit bis Mitte 2018 haben wir die Grundlagenvermittlung in einer Vorlesung kombiniert mit Vorträgen der Studierenden zu aktuellen Fällen. Dabei haben wir die Erfahrung gemacht, dass dieses Format gut geeignet ist, um die Studierenden zu aktivieren. Allerdings zielte dieses Format zu stark auf die reine Vermittlung von Methodenwissen ab. Die Studierenden konnten die Aufgaben mechanistisch abarbeiten und hatten kaum einen Anlass, eine Haltung und Reflexionsfähigkeit zu den Themen zu entwickeln. In der Überarbeitung standen wir also vor der Herausforderung, ein Veranstaltungsformat zu finden, das wesentliche Methoden und aktuelle Themen vermittelt und zugleich Studierende aktiv einbindet und ihnen Möglichkeiten zur Steigerung von Sozialkompetenz, Reflexionsfähigkeit und der Entwicklung einer Haltung/eines Ethos zu entwickeln. Die Forderung nach einer Ethik in der Informatik, die aktuell Eingang in die Diskussion um die Ausbildung von Informatiker*innen findet [GSM21], passt zu diesen Überlegungen.

3 Die ethischen Leitlinien und der ethische Kompass der GI

Die beschriebene Reflexion des Lehrformates I&G an der Fachhochschule Dortmund fällt in die Zeit, in der die Gesellschaft für Informatik (GI) ihre ethischen Leitlinien in einem öffentlich partizipativen Prozess diskutiert, neu formuliert und verabschiedet. Bereits in dem Lehrbuch [KK14] wird die GI als ein wichtiger Repräsentant der Informatik in der Gesellschaft dargestellt und werden die partizipative Ausarbeitung, die Veröffentlichung und Pflege der ethischen Leitlinien als wichtige Aktivitäten der GI dargestellt. In den I&G-Modulen der Fachhochschule Dortmund, die wesentlich auf dem Lehrbuch [KK14] aufgebaut sind, sind die GI und ihre ethischen Leitlinien von Beginn an enthalten.

1994 veröffentlicht die Gesellschaft für Informatik (GI) erstmals ethische Leitlinien, die 2004 in überarbeiteter Form erscheinen. Im Juni 2018 wird die aktuelle Fassung, die aus einem – teilweise digitalen – partizipativen Prozess hervorgeht, vom Präsidium verabschiedet [GIa]. Die Leitlinien bestehen aus einer Präambel, der 12 Artikel folgen. Tabelle 1 nennt die Überschriften der Leitlinien.

Ethische Leitlinien der GI	
1 Fachkompetenz	7 Lehren und Lernen
2 Sachkompetenz und kommunikative Kompetenz	8 Forschung
3 Juristische Kompetenz	9 Zivilcourage
4 Urteilsfähigkeit	10 Soziale Verantwortung
5 Arbeitsbedingungen	11 Ermöglichung der Selbstbestimmung
6 Organisationsstrukturen	12 Die Gesellschaft für Informatik

Tabelle 1: Ethische Leitlinien der GI

Inzwischen gibt es die Leitlinien in zwei Formen: (1) die etablierte Form überschrieben mit „Unsere Ethischen Leitlinien“ [GIb] und (2) die Version mit dem Titel „Ethischer Kompass für Informatik-Fachleute“ [GIc]. Inhaltlich sind beide Varianten identisch. Während sich die klassische Form der ethischen Leitlinien explizit an „GI-Mitglieder“ wendet, richtet sich der ethische Kompass an Informatik-Fachleute, unabhängig davon, ob sie GI-Mitglied sind oder nicht. In der Lehre hat sich herausgestellt, dass Studierende die Einschränkung auf GI-Mitglieder als störend empfinden, weshalb nun die Form des ethischen Kompasses verwendet wird. Im Folgenden werden die Begriffe „ethischer Kompass“ und „GI-Leitlinien“ wechselnd und synonym verwendet.

Die Praxis der I&G Veranstaltungen hat gezeigt, dass die Leitlinien für Studierende eine gute persönliche Ansprache darstellen können. Die Frage „Wie kann ich diese Leitlinie umsetzen?“ trifft die Studierenden sowohl in der Rolle der Anwender*innen von IT-

Systemen als auch in der Rolle der berufstätigen Informatiker*innen, die an der Gestaltung sozio-technischer Systeme beteiligt sind. Aus dieser Beobachtung in einzelnen Terminen der Veranstaltung entstand die Überlegung, das gesamte Lehrkonzept für I&G entlang der ethischen Leitlinien der GI zu organisieren.

4 Lehrkonzept für Bachelorveranstaltung I&G

4.1 Rahmenbedingungen und Überblick

An der Fachhochschule Dortmund ist das Modul I&G Pflicht in den Bachelorstudiengängen Praktische Informatik, Technische Informatik sowie Medizinische Informatik. Jeweils im Wintersemester überschreitet die Zahl der Studierenden im relevanten fünften Fachsemester 150. Im Zuge von Reakkreditierungen werden weitere Studiengänge hinzukommen. Für alle Studiengänge gilt, dass das Modul mit 2,5 ECTS-Punkten gewichtet und unbenotet ist. I&G verwendet das zentrale Lernmanagementsystem der Hochschule ILIAS[3]. Neben der Verwaltung der Kursteilnehmer werden insbesondere Funktionen zur Bereitstellung von Dokumenten, das Buchungswerkzeug für die Gruppeneinteilung und das Diskussionsforum verwendet.

In Abschnitt 2 wird die Herausforderung der Neugestaltung der I&G-Veranstaltung beschrieben, die zeitgleich mit der Neufassung der ethischen Leitlinien der GI auftrat. Da die ethischen Leitlinien bereits ein wichtiges Element der Veranstaltung darstellten und die Studierenden nachweislich ansprachen, entstand ein Lehrkonzept, das diese Leitlinien zur Basis hat. Das Ziel des neuen Lehrkonzeptes ist es, die Sozialkompetenzen der Studierenden stärker zu fördern, sie zu aktivieren und ihnen die Möglichkeit zu geben, eine eigene Haltung zu ethischen Fragen der Informatik auszubilden. Gleichzeitig sollen wesentliche fachliche Elemente von I&G weiterhin vermittelt werden.

Die Kombination unterschiedlicher Lehrmethoden unterstützt diese Ziele. Eine einführende Phase I mit seminaristischen Vorlesungen dient der Organisation und Vermittlung von grundlegendem Fakten- und Methodenwissen. In Phase II werden die ethischen Leitlinien mit aktuellen Fällen aber auch mit Methodenwissen in Verbindung gebracht und von den Studierenden in einem online Forum diskutiert. In der abschließenden Phase III präsentieren und diskutieren die Studierenden in kleinen Gruppen selbst erstellte Poster zu den Leitlinien. Dieser Ansatz orientiert sich an der medienethischen Roadmap aus [Gr18, S. 11]: Ausgehend von der „Sensibilisierung für das Thema“, das durch eine GI-Leitlinie gesetzt wird, folgen die Schritte „Analysieren und Erkennen“, „Auseinandersetzung mit den Auswirkungen“, „Reflexion über die Folgen“, „Wertekonflikte thematisieren“, „Ein Ethos entwickeln“ und

[3] https://www.ilias.de; Stand: 08.07.2022

„Handlungsoptionen aufzeigen“. Die Vermittlung von Methoden- und Konzeptwissen setzt in den Schritten ein, die der initialen Sensibilisierung folgen.

Abb. 1 gibt einen Überblick über die genannten Phasen des Lehrkonzeptes, die in den folgenden Abschnitten weiter erläutert werden.

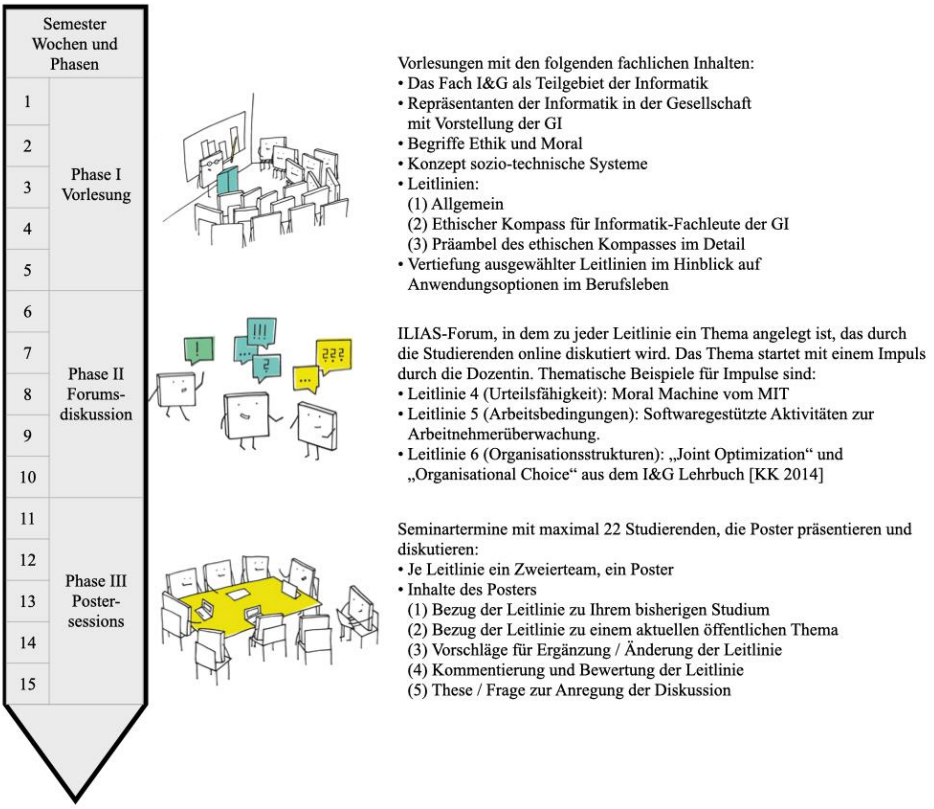


Abb. 1: Überblick Lehrkonzept I&G

4.2 Phase I - Vorlesung

Wie in Abb. 1 dargestellt, beginnt die Veranstaltung I&G mit mehreren Vorlesungen für alle Studierenden der Kohorte. Die Vorlesungen finden wöchentlich statt, dauern 90 Minuten und es gibt keine Anwesenheitspflicht. Diese Phase hat vier Ziele im Hinblick auf die Phasen II und III, die dann vollständig auf Basis des ethischen Fachkompasses organisiert sind. (1) I&G ist als Fach und Teilgebiet der Informatik motiviert und eingeführt. (2) Die GI ist als Autorin des ethischen Fachkompasses bekannt. (3) Die Konzepte Ethik, Moral und dazugehörige Leitlinien im Allgemeinen sind verstanden,

bevor der ethische Fachkompass als ein für die Studierenden relevantes Dokument vorgestellt wird. (4) Einige Leitlinien werden exemplarisch vertieft behandelt, um zu zeigen, welches interdisziplinäre Wissen angewandt werden kann, um die Leitlinie im Berufsleben anzuwenden. Hier wird auch die Anwendung des Konzeptes sozio-technischer Systeme geübt. Phase I erfüllt die Anforderung, für I&G wesentliches Fachwissen zu präsentieren und die Studierenden für die folgenden Phasen vorzubereiten (vgl. auch Abschnitt 2, Forderungen aus [ES12]).

4.3 Phase II - Diskussionsforum

Die Diskussion in dem ILIAS-Forum bildet die zweite Phase der Veranstaltung. Ziel dieser Phase ist, dass sich die Studierenden gemeinsam aktiv mit der praktischen Relevanz der Leitlinien auseinandersetzen. In dem Diskussionsforum legt die Dozentin für jede der Leitlinien ein eigenes Diskussionsthema an. In jedem dieser Diskussionsthemen startet ein Impuls die Diskussion unter den Studierenden. Durch die Auswahl der Impulse sowie weiterer Diskussionsbeiträge platziert die Dozentin Inhalte des Faches I&G, insbesondere aus dem Lehrbuch [KK14]. Beispiele finden sich in Abb. 1.

Die Diskussionsphase dauert ca. fünf Wochen. Die aktive Teilnahme ist für das Bestehen des Moduls Pflicht. Neben der Aufforderung, eine für alle Beteiligten gute Diskussion zu gestalten, erhalten die Studierenden qualitative und quantitative Kriterien, nach denen die aktive Teilnahme bewertet wird. Dazu gehört, dass Beiträge zu mehreren Leitlinien und unterschiedlichen Diskussionssträngen zu leisten sind. So steht in Phase II das in 2 diskutierte Lernziel „argumentativ strukturiertes Reflektieren“ im Mittelpunkt.

4.4 Phase III - Postersessions

Die dritte und letzte Phase der Veranstaltung findet in Form von Seminarterminen mit maximal 22 Studierenden statt. Zweierteams präsentieren und diskutieren nacheinander ihre Poster zu einer gewählten Leitlinie. In jedem Termin werden die Leitlinien 1 bis 11 je maximal einmal vorgestellt[4]. Die Postersession wird für fünf Zeitstunden angesetzt, kann in Abhängigkeit der Teilnehmerzahl aber auch kürzer sein.

Abb. 1 enthält die Vorgaben zu den Postern. Eine Woche vor dem Termin der Postersession müssen die Studierenden ihren Entwurf einreichen und dann ggfs. die Rückmeldungen einarbeiten. Um das Modul I&G zu bestehen, müssen die Studierenden ein Poster präsentieren und aktiv an der gesamten Postersession teilnehmen. In Abgrenzung zu Referaten enthält die Posterpräsentation keine weiteren Folien und ist mit ca. 5 – 10 Minuten reiner Sprechzeit kürzer. Damit bleibt ausreichend Raum für

[4] Da in Leitlinie 12 die GI selbst Akteur ist, wird diese Leitlinie aus Neutralitätsgründen in der Veranstaltung nicht gesondert behandelt.

Diskussionen in der gesamten Gruppe. Ebenso bleibt Zeit für die Dozentin, geeignet Fach- und Methodenwissen auf Basis der von den Studierenden eingebrachten Themen zu vermitteln. So steht in Phase III das Lernziel, eine eigene Haltung zu ethischen Fragen der Informatik zu entwickeln, im Mittelpunkt.

5 Erfahrungen

Das vorgestellte Lehrkonzept wird während der Erstellung dieses Artikels im Sommersemester 2022 zum 8. Mal in Folge an der Fachhochschule Dortmund angewandt. Neben den persönlichen Eindrücken der Dozentinnen stehen weitere Materialien zur Auswertung zur Verfügung: (1) Die hochschulweit standardisierten Lehrevaluationen mit ihren quantitativen und qualitativen Anteilen. (2) Die vollständigen Exporte der Diskussionsforen in HTML-Dateien. (3) Die von den Studierenden erstellten Poster. In diesem Artikel stellen wir einige Schlaglichter dar.

Semester	**Anzahl Beiträge Forum**	**Leitlinien[5] mit den meisten Beiträgen**	**Anzahl Poster-sessions**	**Anzahl bestanden gesamt**
Winter 18/19	750	5 (99) / 4 (79) / 1 (74)	4	65
Sommer 19	492	5 (65) / 4 (53) / 7 (53)	2	44
Winter 19/20	812	5 (127) / 4 (106) / 1 (86)	4	75
Sommer 20	241	9 (28) / 4 (27) / 1 (26)	2	24
Winter 20/21	794	5 (119) / 8 (101) / 7,9 (81)	6	86
Sommer 21	369	4 (60) / 5 (40) / 8 (37)	3	43
Winter 21/22	489	4 (67) / 7 (59) / 5 (55)	5	58

Tabelle 2: Kennzahlen zu I&G-Modulen

Tabelle 2 zeigt einige Kennzahlen zu den I&G-Veranstaltungen seitdem das hier vorgestellte Lehrkonzept im Wintersemester 2018/19 eingeführt worden ist. Die „Anzahl Beiträge im Forum“ ist die Gesamtzahl aller fachlichen Beiträge im Diskussionsforum. Dies beinhaltet sowohl die Beiträge der Dozentin als auch alle Beiträge von Studierenden. Dabei ist zu berücksichtigen, dass wenige Studierende anstelle der Forumsdiskussion eine Alternativaufgabe wählten, in der sie eine eigene gesellschaftlich relevante Tätigkeit in

[5] Die genannten Leitlinien sind: 1 Fachkompetenz, 4 Urteilsfähigkeit, 5 Arbeitsbedingungen, 7 Lehren und Lernen, 8 Forschung, 9 Zivilcourage

den Kontext der Leitlinien stellten. Die Spalte „Leitlinien mit den meisten Beiträgen“ nennt die drei Leitlinien, zu denen die meisten Beiträge geschrieben wurden. In Klammern ist die Zahl der Beiträge zu der jeweiligen Leitlinie angegeben. Die Spalte „Anzahl Postersessions“ nennt die Zahl der Postersessions, die in dem genannten Semester durchgeführt wurden. Die Spalte „Anzahl bestanden gesamt“ nennt die Zahl der Studierenden, die das Modul in dem genannten Semester bestanden haben.

Ein in Abschnitt 2 formuliertes Ziel für eine Lehrveranstaltung I&G ist die aktive Teilnahme der Studierenden. In den Phasen II und III ist dieses Ziel sehr gut erreicht, wobei neben der intrinsischen Motivation der Studierenden auch das Prüfungsformat, das eine Beteiligung zur Pflicht macht, eine Rolle spielt. Die Beiträge der Studierenden sowohl im Diskussionsforum als auch in den Postersessions zeigen, dass es Studierende gibt, die sich nicht nur quantitativ im notwendigen Maße beteiligen, sondern auch inhaltlich hochwertige Beiträge liefern. Anhand der Diskussionsbeiträge und der Poster lässt sich ebenfalls zeigen, dass eine Reflektion der eigenen Rolle als Informatiker*in stattfindet. Auch dies war als Ziel in Abschnitt 2 genannt worden. Der Verlauf einiger Diskussionen belegt, dass die Studierenden auch die Fähigkeit der Argumentation anwenden und einüben.

Neben diesen Beobachtungen aus Dozierendensicht belegen auch Kommentare aus den Evaluationen, dass Studierende das Lehrkonzept mit seinen aktivierenden Anteilen positiv einschätzen, auch wenn die Umsetzung des Diskussionsforums heftig kritisiert wird.

Frage 25: „Was hat Ihnen an dieser Veranstaltung besonders gut gefallen?“
„Lehrkonzept allgemein“ / „Multimediale Lernmethoden“ / „Unterschiedliche Methoden“ /„online-Forum“
Frage 26: „Was könnte künftig besser gemacht werden bzw. wozu hätten Sie sich inhaltlich mehr Bezüge gewünscht?“
„Forum: andere Plattform nutzen, weil wendig übersichtlich. Aber Forumsnutzung wirklich toll. Hat gut auf Postersession vorbereitet. Beitragsmenge war hoch, aber noch angemessen“ „Das Forum ist problematisch: unübersichtlich, keine Benachrichtigung, Emails sind problematisch / zu viele gewesen - technische Umsetzung! Die Idee ist gut!“ „Die Idee einer Diskussion im Forum finde ich besser als eine Vorlesung zu besuchen bei welcher nach 2 Wochen noch 10 Leute teilnehmen würden. Jedoch sollte ein anderes Forum als ILIAS gewählt werden, da sehr unübersichtlich.“ „Verpflichtende Anzahl an Beiträgen innerhalb des Forums sorgt dafür, dass unter anderem viel belangloses geschrieben wird.“

Abb. 2: Beispielhafte Kommentare aus Lehrevaluation

Abb. 2 stellt beispielhafte Fragen und Antworten aus der Lehrevaluation des Wintersemesters 2018/9 dar.

Im Hinblick auf die Anforderung, dass eine I&G-Lehrveranstaltung Bezug auf aktuelle Themen nehmen soll, hat sich gezeigt, dass das Lehrkonzept hierfür ausreichend Raum lässt. Insbesondere in den ersten Impulsen zu jeder Leitlinie in der Forumsdiskussion (vgl. Abb. 1) lassen sich schnell neue Themen einfügen. Aber auch im Verlauf der Diskussion werden aktuelle Themen sowohl von den Studierenden als auch von der Dozentin passend zum Verlauf der Diskussion eingebracht. Die Aufgabenstellung für das Poster enthält explizit den Auftrag, Bezug auf ein aktuelles Thema zu nehmen (vgl. Abb. 1).

6 Diskussion und offene Fragen

Die Erfahrungen zeigen, dass das vorgestellte Lehrkonzept auf Basis der ethischen Leitlinien für Lehrveranstaltungen I&G im Bachelor ein gutes Format ist, um für größere Gruppen von Studierenden sowohl Inhalte zu vermitteln als auch Raum für aktive Beteiligung zu geben. Aktuelle Themen können regelmäßig eingefügt werden, und Studierende erhalten die Möglichkeit zu einem reflektierten argumentativen Austausch, in dem sie ein Ethos im Hinblick auf Anwendung und Gestaltung von IT-Systemen entwickeln können.

Die offenen Fragen beinhalten das Thema einer guten und intensiven Betreuung des Diskussionsforums. Diese ist durch eine Dozentin nicht zeitnah zu der Entstehung der Beiträge zu gewährleisten. Der Kritikpunkt eines Studierenden hinsichtlich der Qualität der Beiträge (vgl. Abb. 2) gibt Anlass, über die Bewertungskriterien für die Forumsteilnahme zu diskutieren. Mit Bezug auf die Postersessions ist die passende Gruppengröße derzeit Gegenstand von Überlegungen.

Um die Inhalte der gesamten Lehrveranstaltung im Workshop zu diskutieren, können kategorisierte Beispiele aus allen Quellen (Evaluation, Diskussionsforum, Postersessions) zur Verfügung gestellt werden.

Auf der Ebene des Selbstverständnisses von I&G als Fach der Informatik (vgl. Abschnitt 1 und 2) geben gegenwärtige Strömungen wie der Begriff der Sozioinformatik [ZKKP21] Anlass zur Diskussion. Die Autorinnen überlegen, eine Neuauflage ihres Lehrbuches [KK14] analog zum Lehrkonzept entlang der ethischen Leitlinien zu organisieren, und stellen diese Überlegung im Workshop zur Diskussion.

Danksagung

Wir bedanken uns bei Prof. Dr. Andreas Harrer, der in konstruktiven Diskussionen zur Entstehung des Lehrkonzeptes beigetragen hat.

Literaturverzeichnis

[DGSW95] Domeyer, V.; Grusdat, M.; Schmidthalks, F.; Wildt, J.: Didaktik von „Informatik und Gesellschaft“. In Friedrich, J.; Herrmann, Th.; Peschke, M.; Rolf, A. (Hrsg.): Informatik und Gesellschaft. Heidelberg: Spektrum Akademischer Verlag, S. 315-321, 1995.

[ES12] Engbring, D.; Selke, H.: Informatik und Gesellschaf als Gebiet der Informatik. In: Forbig, P.; Rick, D.; Schmolitzky, A. (Hrsg.): HDI 2012 – Informatik für eine nachhaltige Zukunft, S. 111-116, 2012. Online: https://publishup.uni-potsdam.de/opus4-ubp/frontdoor/index/index/docId/6252 Stand: 08.07.2022

[GIa] Gesellschaft für Informatik – Ethische Leitlinien. https://gi.de/ueber-uns/organisation/unsere-ethischen-leitlinien Stand: 08.07.2022.

[GIb] Gesellschaft für Informatik – Ethische Leitlinien – Flyer. Online: https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Leitlinien.pdf Stand: 08.07.2022.

[GIc] Gesellschaft für Informatik – Ethischer Kompass. Online: https://gi.de/fileadmin/GI/Allgemein/PDF/GI_Ethischer_Kompass.pdf Stand: 08.07.2022

[Gr18] Grimm, P. et al. (2018). Ethik macht klick – Werte-Navi fürs digitale Leben. Institut für Digitale Ethik (IDE), Hochschule der Medien (HdM) Online: https://www.klicksafe.de/materialien/ethik-macht-klick-werte-navi-fuers-digitale-leben Stand: 08.07.2022

[GSM21] Große-Bölting, G., Scheppach, L., & Mühling, A. (2021). The Place of Ethics in Computer Science Education. In J. Desel, S. Opel, & J. Siegeris (Eds.), Hochschuldidaktik der Informatik - HDI 2021 - 9. Fachtagung des GI-Fachbereichs Informatik und Ausbildung/Didaktik der Informatik. FernUniversität in Hagen, Eigendruck, S. 85-92, 2021.

[HHM98] Herrmann, Th.; Hoffmann, M.; Misch, A. (1998): Wissensvermittlung durch Studierende für Studierende. Stärken und Schwächen einer spezifischen Lehrveranstaltung im Fachgebiet Informatik und Gesellschaft. In: Claus, Volker (Hrsg.): GI-Fachtagung Informatik und Ausbildung, Berlin et al.: Springer. S. 82-91, 1998.

[KK14] Kienle, A.; Kunau, G.: Informatik und Gesellschaft – eine sozio-technische Perspektive. München: de Gruyter, 2014.

[ZKKP21] Zweig, K.; Krafft, T.; Klingel, A.; Park, E.: Sozioinformatik. Ein neuer Blick auf Informatik und Gesellschaft. München: Carl Hanser, 2021.

Human-Centered Data Science – Etablierung einer kritisch-reflexiven Praxis bei der Entwicklung von datengetriebener Software

Claudia Müller-Birn [1] und Lars Sipos[2]

Abstract: Die Erfahrungen der letzten Jahre im Bereich Data Science zeigen immer deutlicher, dass es ein Umdenken bei der Ausbildung bedarf, da die sozialen Nuancen in Daten nicht erfasst oder ethische Kriterien bei der Entwicklung datengetriebener Software zu selten berücksichtigt werden. Diese Lücke soll mit dem Ansatz der Human-Centered Data Science geschlossen werden. Im Rahmen einer Lehrveranstaltung mit Informatikstudierenden haben wir daher diesen Ansatz umgesetzt. Die Ergebnisse einer im Anschluss durchgeführten Interviewstudie zeigen, dass Studierende ein gesteigertes Bewusstsein für die gesellschaftlichen Auswirkungen von Data Science haben. Nichtsdestotrotz hatten wir jedoch den Eindruck, dass sie nicht über eine „Toolbox" verfügen, die es ihnen erlaubt, diese Einsicht nachhaltig in ihren Datenpraktiken zu verankern. Infolgedessen schlagen wir einen konzeptionellen Rahmen für die Förderung einer kritisch-reflexiven Praxis in der Data Science-Ausbildung vor, den wir in diesem Artikel näher vorstellen. Anhand einer beispielhaften Umsetzung erläutern wir jede der vier aufeinander aufbauenden Phasen, die in zukünftiger Forschung evaluiert werden.

Keywords: Human-Centered Data Science; Ethik

1 Die Motivation für Human-Centered Data Science

Das Forschungsgebiet Data Science hat sich in den letzten Jahren rasant entwickelt und dies stellt die akademische Ausbildung vor besondere (wenn auch nicht neue) Herausforderungen (siehe [Va17]). Die Erfahrungen der letzten Jahre zeigen immer deutlicher, dass der Fokus auf statistische und numerische Aspekte in der Ausbildung nicht ausreicht, um soziale Nuancen, affektive Beziehungen, ethische, werteorientierte Grundsätze oder die tatsächlichen Auswirkungen einer datengetriebenen Software auf die Gesellschaft zu erfassen [BA17, Ko20, Ar22]. Aus diesen Herausforderungen hat sich das Gebiet *Human-Centered Data Science* entwickelt, in welchem die komplexen Interaktionen zwischen Gesellschaft, Technologie und von Menschen erzeugten Daten erforscht werden [Ar22]. Im Mittelpunkt steht dabei, die herkömmlichen computergestützten Methoden zur Analyse großer Datensätze mit qualitativen Methoden zu verbinden, die mit ihrem Detailreichtum und Kontextwissen zu einem tieferen Verständnis von Daten und Gesellschaft beitragen

[1] Freie Universität Berlin, Institut für Informatik, Human-Centered Computing, clmb@fu-berlin.de, https://orcid.org/ 0000-0002-5143-1770
[2] Freie Universität Berlin, Institut für Informatik, Human-Centered Computing, lars.sipos@fu-berlin.de

doi:10.18420/inf2022_62

können. Ein zentrales Anliegen in der Human-Centered Data Science ist es, den Studierenden eine kritisch-reflexive Datenpraxis zu vermitteln. Dieser Artikel soll einen Diskussionsbeitrag liefern, wie eine solche kritisch-reflexive Datenpraxis in der akademischen Ausbildung im Bereich Data Science verankert werden kann. Es wird dazu ein laufendes Lehrangebot skizziert und erste Erkenntnisse aus der Umsetzung vorgestellt.

Human-Centered Data Science ist ein interdisziplinäres Forschungsgebiet, das sich auf Erkenntnisse und Methoden aus den Bereichen der Mensch-Computer-Interaktion, den Sozialwissenschaften, der Statistik und des maschinellen Lernens stützt [Ar22]. Der Human-Centered Data Science liegt dabei ein menschenzentrierter Ansatz bei der Technologiegestaltung - das Human-Centered Design - zugrunde, um die Praktiken der Data Science zu verbessern. Dieser Ansatz der menschenzentrierten Gestaltung basiert auf einer Reihe von Leitsätzen, die Kling & Star bereits vor über 20 Jahren aufgestellt haben [KS98]. Danach sollte datengetriebene Software menschliche Fähigkeiten sinnvoll ergänzen, aber diese nicht ersetzen oder automatisieren. Soziale Konstrukte (z.B. Fairness) sollten nicht in mathematische Konzepte übersetzt werden, da mathematische Operationalisierungen die Vielfältigkeit der sozialen Realität nicht umfassend zu beschreiben vermögen. Somit sollte bei der Entwicklung datengetriebener Software nicht nur die Optimierung der statistischen Modelle im Mittelpunkt stehen, sondern auch der Kontext berücksichtigt werden, in dem die Software letztlich eingebettet ist. Dies erfordert es, dass auch Fragen der ökologischen Nachhaltigkeit berücksichtigt werden. Kling & Star verweisen bereits darauf, dass Entwickler:innen Bescheidenheit („modesty“) in Bezug auf die Fähigkeiten von datengetriebener Software entwickeln sollten, denn Technologie allein kann keine sozialen Probleme, wie beispielsweise die der Gerechtigkeit, lösen. Eine wesentliche Grundlage der menschenzentrierten Gestaltung bildet daher ein Verständnis über die in die datengetriebene Software inkludierten Werte, die der zukünftigen direkt oder indirekt durch den Softwareeinsatz betroffenen Personengruppen, aber auch die der Datenwissenschaftler:innen selbst. Durch eine solche menschenzentrierte Perspektive auf Data Science können eine Vielzahl von Methoden aus dem Bereich der Mensch-Computer Interaktion in der Data Science verankert werden, wie beispielsweise die der partizipativen Gestaltung. Eine grundlegende Voraussetzung in der Vermittlung dieser Methoden ist es aber, die zumeist positivistisch geprägte Haltung von Studierenden um eine kritisch-reflexive Datenpraxis zu bereichern.

2 Erste Umsetzung und Interviewstudie

Ausgehend von diesen Überlegungen haben wir im Wintersemester 2020/21 erstmalig die Veranstaltung Human-Centered Data Science für Studierende der Informatik und der Data Science angeboten, um sie mit dieser Perspektive vertraut zu machen. Aktuell werden bereits Veranstaltungen angeboten, die die Human-Centered Design-Perspektive einbeziehen (z. B. in den Bereichen Mensch-Computer-Interaktion, Datenvisualisierung) sowie Veranstaltungen im Bereich des Maschinellen Lernens. Wir haben jedoch festgestellt, dass

ein Kurs, der die gesellschaftlichen Auswirkungen von bestehenden Datenpraktiken hervorhebt, in der derzeitigen Ausbildung bisher noch fehlt.
Die Veranstaltung hatte eine typische Struktur und bestand aus einer Vorlesung (2 SWS), in der die Konzepte, Techniken und Ansätze theoretisch eingeführt wurden, und einer Übung (2 SWS), in der ausgewählte Themen praktisch angewandt wurden. Die Themen wurden primär aus den bestehenden Diskursen der ACM Conference on Human Factors in Computing Systems (CHI) und der ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) Community abgeleitet.[3] Eine disziplinär heterogene Gruppe von Studierenden (vier Frauen und zehn Männer) nahm an der Veranstaltung teil: neben Informatik und Data Science gab es Studierende der Wirtschaftsinformatik, Ergonomie und Politikwissenschaft.

2.1 Studienaufbau

Nach der ersten Durchführung der Veranstaltung waren wir an den Eindrücken der Studierenden interessiert und führten daher eine Interviewstudie durch. Wichtig war dabei, Studierenden die Möglichkeit zu bieten, ihre persönlichen Erfahrungen direkt mit uns zu teilen und die Effektivität unseres Lehransatzes, bei der Vermittlung einer Human-Centered Data Science zu bewerten. Wir luden alle Studierenden ein, an einem Interview teilzunehmen, wovon sechs Studierende der Einladung folgten. Die Interviews folgten einem semi-strukturiertem Format, um genügend Raum für Diskussionen über unerwartete Aspekte zu schaffen. In den Interviews fragten wir beispielsweise Studierende, ob sie das Gefühl hatten, die Lernziele erreicht zu haben, wie sie die allgemeine Struktur der Veranstaltung und das Tempo erlebten und stellten Fragen zu den verwendeten Tools und Methoden. Nach der Transkription der Interviews führten wir eine Thematische Analyse durch. Dabei folgten wir dem Ansatz von Braun & Clarke [BC06] aufgrund ihrer Flexibilität, Zugänglichkeit und ihres Potenzials, unerwartete Erkenntnisse zu gewinnen. Wir analysierten die Daten vor allem dahingehend, inwieweit es uns gelungen ist, das kritische Verständnis der Studierenden für die unterrichteten Themen zu fördern. Die Transkripte wurden in Bezug auf die Lernziele kodiert und alle Codes und Zitate wurden schließlich zur Generierung der Themen zusammengetragen.[4]

2.2 Studienergebnisse

Innerhalb des Analyseprozesses konnten wir eine Reihe von Themen identifizieren, wovon wir nachfolgend drei zentrale vorstellen: „Erhöhtes Bewusstsein für die gesellschaftlichen Auswirkungen von Data Science“, „Zwiespältiges Interesse an den Themen des

[3] Eine Übersicht zu den Themen gibt es auf der Veranstaltungsseite: https://www.mi.fu-berlin.de/en/inf/groups/hcc/teaching/Past-Terms/winter_term_2020_21/course_human_centered_data_science.html

[4] Einen Überblick über die Fragen und die Transkripte teilen wir auf OSF: https://doi.org/10.17605/OSF.IO/JD6TQ

Kurses und seinem Bezug zur Informatik“ und „Beispiele aus der Praxis und Reflexion unterstützten das Verständnis der Studierenden“.
Obwohl die Studierende keine homogene Gruppe darstellen und eine Vielzahl unterschiedlicher Hintergründe und Vorerfahrungen besitzen, berichteten alle interviewten Studierenden von einem gesteigerten Bewusstsein dafür, wie sich aktuelle Entwicklungen in der Data Science auf die Gesellschaft auswirken. In einem Auszug, der dies betont, wird erwähnt, dass diese Veranstaltung den Studierenden die „Werkzeuge“ gab, um über diese Entwicklungen zu sprechen: *„[…] ich muss sagen, dass ich auch vorher schon in diese Richtung aufgepasst hab und auch eine ‚Awareness‘, also ein Bewusstsein dafür hatte, aber vielleicht jetzt noch nicht die Tools hatte. Es war eher ein diffuseres ‚Da muss man aber aufpassen!‘ und ich kannte auch schon ein paar Beispiele, wo eben Dinge passiert sind, die nicht so super sind […] aber ich glaube, was mir auf jeden Fall geholfen hat ist, dass man bestimmte Sachen einfach besser definieren, oder bezeichnen kann, also Worte für bestimmte Dinge hat, wo man sagen kann so und so*“ (S4). Studierende hatten also bereits ein latentes Gefühl für bestimmte Themen, sei es durch frühere Erfahrungen, Nachrichtenartikel oder ähnliches. Sie verfügten jedoch nicht über die richtigen Konzepte und Werkzeuge, um das Ausmaß der mit diesen Technologien verbundenen Probleme vollständig zu erfassen und auch nicht über die Sprache, um diese Themen mit ihren Kommiliton:innen oder potenziellen Interessengruppen zu diskutieren. Für das kritische Denken ist es jedoch unabdingbar, das richtige Vokabular zu besitzen, um solche Gedanken ausformulieren zu können. Die Teilnehmer:innen konnten am Ende teilweise sich selbst kritische Fragen zu ihren Arbeitsweisen stellen: „*Okay, wo können Fehler entstehen?*“ (S4), „*Ist das [System] überhaupt gut, in jedem Zusammenhang einzusetzen?*“ (S3). In Bezug auf die Existenz von Bias hat eine Person gut in Worte gefasst, dass sie nun weiß, „*[…] wie man [die Daten] analysiert, überwacht, vorbereitet und sich ihnen kritisch entgegenstellen kann*“ (S4). Darüber hinaus sind sich die Studierenden nun bewusster, wie wichtig es ist, Blackbox-Modelle und Bias in den Daten interpretieren zu können: „*Ich sehe schon einen Wert darin, dass man auf jeden Fall eine Blackbox aufmachen und ein wenig reinschauen kann*“ (S2). Weiterhin können sie sich auf einen menschenzentrierten Ansatz besinnen und die reale Welt sowie die Menschen, welche datengetriebene Systeme nutzen oder von ihnen betroffen sind, berücksichtigen: „*[…] praktisch hat das [System] in der realen Welt Leute beeinflusst und Auswirkungen auf sie gehabt und zwar ständig*“ (S3).

In Bezug auf die eigene Arbeit wurde festgestellt, dass die entscheidenden Anstrengungen vor der Implementierung und Programmierung der Systeme unternommen werden: „*[…] wenn ich jetzt daran [einem Data Science Projekt] sitzen würde, dass ich nicht nur ans Programmieren denke und wie kriege ich irgendwas umgesetzt, sondern dass das [die Analyse] nicht nebensächlich ist […] ja [man überlegt], was man priorisieren sollte und das ist definitiv nicht nur das Programmieren*“ (S4). Weiterhin wurde die Erkenntnis erlangt, dass Fairness für Informatiker:innen wichtig ist sowie dass der Einsatz von nicht fairen Modellen des maschinellen Lernens sogar nutzlos ist: „*Ich wusste nicht, warum ist das jetzt wichtig auf Fairness zu achten? ‚Ach Gott, dann ist ja jetzt diese eine Maschine ein bisschen diskriminierend‘ ist ja jetzt besser als gar nichts zu haben vielleicht. Aber dann hat es bei mir klick gemacht. Ja ne, eine falsche Maschine, das ist falsch‘. Es ist ja*

nicht nur ein bisschen falsch, es ist, wenn man es so sieht, einfach nur falsch und das kann man dann später für die Produktion oder so gar nicht benutzen“ (S2).

Es gab zwei eingesetzte Methoden, die von den Befragten als besonders hilfreich empfunden wurden, sowie eine, die auf weniger Resonanz stieß. Die meisten Studierenden verinnerlichten besonders Praxisbeispiele, die Probleme in realen sozio-technischen Kontexten beschrieben. Diese Ansicht wurde in allen Interviews geteilt, wie in diesem: „*[…] was ich immer schön fand war, dass man Praxisbeispiele hatte. Das fand ich immer sehr cool. Meistens waren das immer Sachen, die schief gelaufen sind und überhaupt nicht gut waren. Ich fand das, irgendwie sehr cool zu sehen, was dann auch passieren kann und was dann eben schon in der Welt passiert ist an absurden Dingen, die mit Data Science gemacht werden und welche sozialen Auswirkungen das auf Menschen hat*“ (S4). Überraschenderweise wurden Lehrmethoden, welche die Studierenden als aus anderen Disziplinen stammend ansahen, „*Ich weiß auch von anderen Studiengängen, dass da ganz viel reflektiert wird; immer über irgendwelche Hausaufgaben und alles Mögliche. Das ist da ein häufiger Aufgabenbestandteil*“ (S3), als wirklich hilfreich angesehen, da die Studierenden bemerkten: „*[…] okay, so eine Reflektion bringt wirklich was*“ (S6). Eine Person beschrieb, obwohl sie zuvor erwähnt hatte „*Programmieren mag ich immer mehr*“ (S4), dass das Aufschreiben ihr bei der Reflektion von Themen wirklich geholfen hat: „*[…] obwohl ich schreiben nicht ganz so gerne mag, aber es halt sein muss und man halt dadurch auch besser wird und irgendwann merke ich langsam auch, dass ich an einem Punkt bin, dass ich dann doch schreiben ganz gut finde und auch sehr sehr wichtig, weil man dabei auch viel reflektiert und beim Coding eigentlich nicht viel reflektieren muss zum Teil. Hauptsache es funktioniert halt*“ (S4). Wenig überraschend war, dass die große thematische Breite für die Teilnehmer:innen eine Herausforderung darstellte. Es war eine komplexe und schwierige Aufgabe, all die verschiedenen Aspekte des HCDS in einem einzigen Kurs zu behandeln. Insbesondere das Thema Bias wurde von den Studierenden wiederholt als Beispiel für diese Komplexität angesprochen. Häufig wurde beklagt, dass in der Vorlesung zu viele verschiedene Bias Formen behandelt wurden und dass die Studierenden nach einer Weile einfach den Anschluss verloren: „*[…] es ist so schwer für mich, der Vorlesung zu folgen, weil ich zum ersten Mal so viele verschiedene Arten von Vorurteilen höre*“ (S1).

2.3 Einordnung der Ergebnisse

Die erste Version des Kurses wurde aufgrund mangelnder bestehender Lehrbeispiele und nicht bestehender Textbücher recht intuitiv entworfen, wobei wir uns weitgehend von der bestehenden Forschung und unseren eigenen Interessen leiten ließen. Aufgrund unserer eigenen Erfahrungen und der Ergebnisse der Interviewstudie haben wir uns jedoch entschlossen, genauer über unseren Vermittlungsansatz nachzudenken.

Bei der Erörterung des Feedbacks der Studierenden stellte sich heraus, dass das größte Bedürfnis darin bestand, die Werte und grundlegenden Thesen zu vermitteln, die das Verständnis und die Problemlösung von Informatiker:innen in Human-Centered Data Science-Projekten prägen (vgl. [OV21]). Wir stellten fest, dass wir zwar die verschiedenen

Methoden, Konzepte und Techniken, die zugrundeliegenden Werte und grundlegenden Thesen, die die Praxis der Human-Centered Data Science prägen, vorgestellt hatten, diese aber nicht ausreichend an praktischen Kenntnissen und Fähigkeiten ausgerichtet hatten. Infolgedessen hatten die Studierenden zwar das Gefühl, ein Vokabular für die Diskussion bestehender Probleme in ihrer Data-Science-Praxis gelernt zu haben, wir hatten aber den Eindruck, dass sie nicht über einen ausreichenden „Werkzeugkasten" verfügten, diese Probleme in ihrer Praxis dann auch nachhaltig anzugehen.

Dieser „Werkzeugkasten" sollte gerade nicht aus Metriken, Techniken, Algorithmen oder Software bestehen, sondern aus einer "Haltung" bzw. "Einstellung", die es Studierenden erlaubt, über diese Probleme nachzudenken. Nachdem wir diesen Eindruck diskutiert hatten, wollten wir das Bewusstsein für eine ethische Verantwortung in Human-Centered Data Science enger mit der Praxis verknüpfen. Daher haben wir uns insbesondere mit Lehransätzen aus dem Bereich der Ethik auseinandergesetzt, wie wir nachfolgend vorstellen werden.

3 Förderung kritisch-reflexiver Praktiken in der Data Science

Eine Erkenntnis aus den Herausforderungen des Einsatzes von datengetriebener Software im gesellschaftlichen Kontext war, dass Informatiker:innen oder Datenwissenschaftler:innen neben der Vermittlung von technischen Fähigkeiten auch im ethischen Denken geschult werden sollten. Bestehende Ethik-Lehrveranstaltungen verfolgen das Ziel, den Studierenden beizubringen, ethische Probleme in der Welt zu erkennen, diese Probleme kritisch zu beurteilen und Technologien in Bezug auf diese Probleme zu bewerten sowie gut begründete Argumente auf der Grundlage der Kritik zu formulieren [FGB20]. Ethik wird in diesen Veranstaltungen aber eher selten aus der Entwicklungspraxis heraus operationalisiert. Es werden vielmehr ethische Theorien (z.B. Utilitarismus, Deontologie) erläutert und anhand moralphilosophischer Gedankenexperimente (z.B. Trolley-Problem) diskutiert. Es werden darüber hinaus bestehende ethische Grundsätze (z.B. Privatsphäre, Rechenschaftspflicht, Transparenz und Erklärbarkeit, Fairness und Nichtdiskriminierung) eingeführt, aber deren konkrete Umsetzung in der jeweiligen Data Science-Praxis bleibt unterspezifiziert. Dies hat damit zu tun, dass die Ethik-Ausbildung in fachlich losgelöste Veranstaltungen ausgelagert und häufig aus Sicht der Geistes- oder Sozialwissenschaft vermittelt wird. Dies ist fachlich sicherlich sinnvoll, aber damit ethische Überlegungen auch über den jeweiligen Lehrkontext hinaus nachhaltig wirken können, sollten diese stattdessen in Data Science-Praktiken eingebettet sein.

Es bedarf in der Data Science-Ausbildung einer Handlungsethik (Ethics-in-Action), wie sie von Frauenberger et al. formuliert wurde [FRF17]. Die Handlungsethik ergänzt dabei den durch bestehende ethische Grundsätze (formuliert durch Regierungen, Unternehmen, Fachgesellschaften und Nichtregierungsorganisationen) formalisierten institutionalisierten Ethik-Rahmen, verbindet diesen aber gleichzeitig mit der praktischen Datenarbeit. Die Handlungsethik baut auf Donald Schöns Reflective Practice auf, nach dem eine Person

durch einen gewählten methodischen Ansatz in einen Dialog mit einer Situation tritt. Dieser Dialog „*should create a dynamic in which the situation 'talks back', and [the designer] responds to the situation's back-talk*" [Sc83]. Schön bezeichnet dieses Gespräch mit einer Situation als reflexiv [Sc83]. Darüber hinaus unterscheidet Schön zwischen einer Reflection-in-Action (Handlungsreflexion) und Reflection-on-Action (Reflexion über die Handlung). Erstere findet während der Handlung statt und ergibt sich aus unerwarteten Ergebnissen einer Handlung. Letztere beschreibt eine Reflexion, die auf der Grundlage einer früheren Handlung erfolgt und mit einer Phase der Rückschau beginnt, in der diejenigen, die reflektieren, die frühere Handlung überprüfen. Im Bereich Data Science kann eine solche „Situation", in der eine Handlungsreflexion erfolgt, als konkrete Entscheidung bei der Datenauswahl, der Datenaufbereitung, dem Feature Engineering etc. angesehen werden. Bei diesen Entscheidungen ist es wichtig, diesen „Dialog" zur Handlungsreflexion herbeizuführen, der letztlich Auswirkungen auf die Entscheidungen hat.

Aus der Handlungsethik und der Handlungsreflexion ergibt sich somit die Notwendigkeit, ethische Überlegungen eng in die Daten- und Programmierpraxis von Datenwissenschaftler:innen einzubetten. Der Dialog kann nur durch bestehende ethische Grundsätze (z.B. Fairness und Nichtdiskriminierung) initiiert werden, aber das reflexive Handeln selbst muss zu einem inhärenten Teil der Data Science-Praxis werden. Um diese Verbindung herzustellen, wird ēthos (altgriechisch „Charakter") benötigt, "*a moral commitment or stance, a moral attitude that underlies a particular practice*" [FRF17]. Im Gegensatz zu formalen Ethikgrundsätzen ist ēthos also intrinsisch und personifiziert. Somit ist bei der akademischen Ausbildung ein wichtiges Ziel, dass sich Studierende ihrer Werte bewusst werden. Diese Werte bilden letztlich die Grundlage für die Anwendung der ethischen Grundsätze. Es ist somit von zentraler Bedeutung, dass Studierende diese eigenen Werte formulieren und danach handeln können.

Basierend auf diesen Überlegungen stellen wir nachfolgend unseren Vorschlag für einen methodischen Ansatz für die Vermittlung von Human-Centered Data Science in der akademischen Ausbildung vor, welche das Ziel verfolgt, die kritisch-reflexive Praxis der Data Science-Studierenden zu fördern, indem die technischen und ethischen Kompetenzen gestärkt werden. Wir verwenden dazu bestehende ethische Grundsätze (z.B. Fairness und Nichtdiskriminierung, Transparenz und Erklärbarkeit) und verknüpfen sie mit konkreten Implementierungsaufgaben. Jeder ethische Grundsatz wird anhand einer Fallstudie eingeführt, um Interesse zu wecken. Dieser Fall wird durch Studierende anhand ihrer Werte reflektiert, um potenzielle Lösungsansätze abzuleiten, die dann anhand verfügbarer "Lösungsansätze" umgesetzt und praktisch evaluiert werden. Als konzeptioneller Rahmen dienen dabei Greens vier Phasen einer Critical Reflexive Technical Practice [Gr21]. Nachfolgend erläutern wir diese vier Phasen (Interesse, Reflexion, Anwendungen und Praxis) exemplarisch anhand des ethischen Grundsatzes der Nichtdiskriminierung.

Die ersten zwei Phasen werden in der Vorlesung durchlaufen, während die letzten beiden Phasen Teil der Übung sind. Diese Phasen können innerhalb einer Lehrveranstaltung mehrmals unter Anwendung verschiedener ethischer Grundsätze durchlaufen werden.

4 Beispiel einer kritisch-reflexiven Praxis anhand des Grundsatzes der Nichtdiskriminierung

In der ersten Phase geht es vor allem darum, Interesse für das Problem (abgeleitet aus dem ethischen Grundsatz) zu wecken. Dazu sollten gesellschaftlich relevante Daten anstelle von banalen Beispieldaten verwendet und konkrete gesellschaftliche Probleme diskutiert werden. Dies zielt darauf ab, das Denken der Studierenden vom einfachen Technologieeinsatz auf die positive Beeinflussung der Gesellschaft zu lenken [Gr21]. Hierfür verwenden wir Fallstudien, die eine bestimmte „reale" Situation beschreiben. Die Fallstudien dienen dazu, Studierende nach erfolgtem theoretischem Input (z.B. im Bereich der Diskriminierung) zu aktivieren. Geeignete Beispiele finden sich in der Sammlung der Gewissensbits [We09], die von Mitgliedern der Fachgruppe „Informatik und Ethik" der Gesellschaft für Informatik in den letzten zehn Jahren erstellt wurden[5]. Die Fallstudien decken ein breites Spektrum gesellschaftlicher Themen ab, wodurch Studierende leichter nachvollziehen können, inwieweit eine bestimmte Technologie auch ihr unmittelbares Leben beeinflussen kann.

In der zweiten Phase soll die Reflexion gefördert werden, indem Erkenntnisse und Theorien aus Bereichen wie der Wissenschafts- und Technikforschung, Technikphilosophie, Soziologie und Politikwissenschaft bei der Beurteilung einer „Situation" einbezogen werden. Die Situation wird dabei in Daten, Modell, Zielgruppe, Zweck, Kontext etc. zerlegt, während die erörterten Theorien dazu anregen sollen, über bestehende Annahmen in diesen Bereichen nachzudenken. Durch die Diskussion der Fallstudie wird es Studierenden ermöglicht, über ethische Herausforderungen in der Data Science nachzudenken und zu erkennen, dass es oft keine eindeutigen Lösungen gibt. Ziel ist es, Studierende zu befähigen, die eigenen Argumente nachvollziehbar zu formulieren und darzustellen sowie gegenteilige Argumente aufzunehmen und zu bewerten. Ein wesentlicher Ansatz in der Diskussion ist es daher, auf andere Meinungen einzugehen und gemeinsame Lösungen zu suchen. Eine wichtige Erkenntnis in diesen Diskussionen soll es sein, dass bestehende Zielkonflikte beispielsweise bezüglich des Ressourcenverbrauchs, des Sicherheitsniveaus und der Zuverlässigkeit nicht eindeutig lösbar sind. Darüber hinaus sollen Studierende verstehen, dass datengetriebene Software innerhalb eines soziotechnischen Kontexts eingebettet ist und Zielkonflikte auch dadurch häufig nicht lösbar sind, da es nicht nur um technische Lösungen geht.

In der dritten Phase konzentrieren wir uns darauf, Studierenden dabei zu unterstützen, die erlernten Konzepte/Perspektiven in ihre Data Science-Praxis einzubringen. Daher stellen wir Studierenden Datensätze zur Verfügung (z.B. Kreditdaten[6]), welche sie zur Umsetzung einer konkreten Anwendung (z.B. Kreditberatungssoftware) verwenden sollen. Im

[5] Gewissensbits werden regelmäßig in einer Rubrik der Fachzeitschrift Informatik Spektrum von Springer veröffentlicht. Weitere Informationen finden Sie unter http://gewissensbits.gi.de

[6] Der Datensatz ist verfügbar unter https://archive.ics.uci.edu/ml/datasets/Statlog+\%28German+Credit+Data\%29

Themenbereich der Diskriminierung erarbeiten Studierende durch den Einsatz der „Datasheets for Datasets“ [Ge21], was sie über den Datensatz wissen und was nicht. Wir diskutieren anschließend, wie durch diese Dokumentation die Nutzbarkeit solcher Datensätze erhöht werden könnte. Anschließend untersuchen die Studierenden mögliche Verzerrungen im Datensatz mithilfe des AI360 Toolkit[7]. Das Ziel dieser praktischen Übungen ist es immer wieder, dass Studierende lernen, ihre Entscheidungen zu hinterfragen, um so eine kritische Denkweise aufzubauen. Dabei geht es auch darum, bestehende Techniken, Ansätze und Hilfsmittel zu kritisieren, d.h. deren Grenzen zu verstehen und Veränderungsvorschläge zu erarbeiten.

In der letzten Phase sollen Studierenden in die Lage versetzt werden, eine Human-Centered Data Science-Praxis zu realisieren. Dabei lernen Studierende Methoden aus der menschenzentrierten Gestaltung kennen und wenden diese innerhalb eines praktischen Beispiels an. Beispielsweise geht es bei der Kreditberatungssoftware nicht nur darum, dass Studierende ein entsprechendes Vorhersagemodell entwickeln, sondern auch mit den potenziellen Gruppen, die vom Einsatz dieser Anwendung (direkt oder indirekt) betroffen sind, in Kontakt treten und die Auswirkungen ihrer Software verstehen lernen. Dazu werden zunächst Evaluationsstudien durchgeführt, indem die Studierenden, die in den vorhergehenden Phasen realisierten Anwendungen (vor allem in Form von Jupyter Notebooks) kennenlernen und gegenseitig beurteilen. Durch die gegenseitige Evaluation der unterschiedlichen Anwendungen werden die Studierenden ermutigt, die „beste“ Anwendung zu küren. Auch diese Aufgabe soll Studierende befähigen, verschiedene Perspektiven auf das Problemfeld einzunehmen und auch potenzielle Konsequenzen zu antizipieren.

5 Zusammenfassung und Ausblick

Human-Centered Data Science ist ein interdisziplinäres Forschungsgebiet, welches sich vor allem auf einen menschenzentrierten Ansatz der Technologiegestaltung fokussiert und dabei Erkenntnisse aus den Bereichen Mensch-Computer-Interaktion, Sozialwissenschaften, Statistik und maschinelles Lernen integriert. Bestehende Lehrkonzepte zur Vermittlung dieses Forschungsgebiet existieren bisher noch nicht. Mit dem vorliegenden Artikel möchten wir diese Lücke schließen und bauen auf Schöns Handlungsreflexion [Sc83] und Greens Phasen einer Critical Reflexive Technical Practice [Gr21] auf, um eine kritisch-reflexive Praxis von Studierenden zu fördern.

Aktuell wird das hier vorgestellte Lehrkonzept praktisch in einer laufenden Lehrveranstaltung evaluiert. Für die Zukunft planen wir, alle Materialien als frei zugängliches Lehrmaterial unter einer offenen Lizenz zur Verfügung zu stellen.

Danksagung

Die Autor:innen bedanken sich für die finanzielle Unterstützung beim Bundesministerium für Bildung und Forschung (BMBF) im Rahmen des Vorhaben "ENKIS – Etablierung

[7] Weitere Informationen zu dieser Software sind unter https://github.com/Trusted-AI/AIF360 verfügbar.

nachhaltiger KI-bezogener Studienangebote für eine Verantwortungsvolle Künstliche Intelligenz an der Freien Universität Berlin" (Förderkennzeichen: 16DHBKI025) und bei allen Reviewer:innen für die hilfreichen Verbesserungsvorschläge.

Literaturverzeichnis

[Ar22] Aragon, Cecilia; Guha, Shion; Kogan, Marina; Muller, Michael; Neff, Gina: Human-Centered Data Science: An Introduction. MIT Press, Cambridge, MA, USA, 2022.

[Ba17] Baumer, Eric P. S.: Toward human-centered algorithm design. Big Data & Society, 4(2):2053951717718854, 2017.

[BC06] Braun, Virginia; Clarke, Victoria: Using Thematic Analysis in Psychology. Qualitative Research in Psychology, 3(2):77–101, 2006.

[FGB20] Fiesler, Casey; Garrett, Natalie; Beard, Nathan: What Do We Teach When We Teach Tech Ethics? A Syllabi Analysis. In: Proceedings of the 51st ACM Technical Symposium on Computer Science Education. Association for Computing Machinery, New York, NY, USA, S. 289–295, 2020.

[FRF17] Frauenberger, Christopher; Rauhala, Marjo; Fitzpatrick, Geraldine: In-Action Ethics. Interacting with Computers, 29(2):220–236, Mar 2017.

[Ge21] Gebru, Timnit; Morgenstern, Jamie; Vecchione, Briana; Vaughan, Jennifer Wortman; Wallach, Hanna; III, Hal Daumé; Crawford, Kate: Datasheets for datasets. Communications of the ACM, 64(12):86–92, Nov 2021.

[Gr21] Green, Ben: Data Science as Political Action: Grounding Data Science in a Politics of Justice. Journal of Social Computing, 2(3):249–265, Sep 2021.

[Ko20] Kogan, Marina; Halfaker, Aaron; Guha, Shion; Aragon, Cecilia; Muller, Michael; Geiger, Stuart: Mapping Out Human-Centered Data Science: Methods, Approaches, and Best Practices. In: Companion of the 2020 ACM International Conference on Supporting Group Work. GROUP '20, ACM, New York, NY, USA, S. 151–156, 2020.

[KS98] Kling, Rob; Star, Susan Leigh: Human Centered Systems in the Perspective of Organizational and Social Informatics. SIGCAS Comput. Soc., 28(1):22–29, 1998.

[OV21] Olsson, Thomas; Väänänen, Kaisa: How Does AI Challenge Design Practice? Interactions, 28(4):62–64, 2021.

[Sc83] Schön, Donald A.: The reflective practitioner: how professionals think in action. Basic Books, New York, 1983.

[Va17] Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N; Kaiser, Łukasz; Polosukhin, Illia: Attention is all you need. Advances in neural information processing systems, 30, 2017.

[We09] Weber-Wulff, Debora; Class, Christina; Coy, Wolfgang; Kurz, Constanze; Zellhöfer, David: Gewissensbisse: Ethische Probleme der Informatik. Biometrie - Datenschutz - geistiges Eigentum. transcript Verlag, 2009.

Ethische Fragen und informatische Modelle – eine symbiotische Beziehung in der informatischen Bildung

Eine ethische und bildungstheoretische Perspektive auf die politische Dimension von Technik

Hermann Diebel-Fischer [1] und Lutz Hellmig[2]

Abstract: Dieser Beitrag wirft aus technikethischer Perspektive ein Schlaglicht auf (vermeintliche) Partizipationsmöglichkeiten in der Technikentwicklung, die vor dem Hintergrund der gesellschaftlichen Bedeutung bestimmter Technologien gefordert werden. In diesem Zusammenhang werden strukturelle Bedingungen aufgezeigt, die Beteiligungsformate in ihrer Wirksamkeit beschränken und es werden Implikationen für die technische, insbesondere die informatische Bildung im schulischen wie gesellschaftlichen Kontext aufgezeigt. Um eine Grundlage für eine sinnvolle Partizipation in der technischen Gestaltung der Gesellschaft zu schaffen, wird für den Prozessbereich des Modellierens und Implementierens der GI-Bildungsstandards ein Erweiterungsvorschlag gemacht.

Keywords: Ethik, informatische Bildung, Modellierung, GI-Bildungsstandards, Partizipation

1 Technik als politische Struktur

Wenn ein Textverarbeitungsprogramm Vorschläge zur besseren Lesbarkeit eines Textes macht oder man am Flughafen einer gesonderten Sicherheitskontrolle unterzogen wird, dann liegt vordergründig das Arbeitsergebnis von Informationstechnik vor. So unterschiedlich die Situationen sein mögen, so ähnlich ist doch ihr Zustandekommen mit Blick darauf, wie das Verhältnis von Technik und Betroffenen konfiguriert ist: Über die Einführung von Technologien in die Gesellschaft wird diese nicht nur dahingehend verändert, dass weitere Werkzeuge zur Verfügung stehen (etwa um besser lesbare Texte zu produzieren oder die Gefahr krimineller Handlungen zu senken), sondern auch insofern, als diese Technologien als Ergebnisse von (teilweise verdeckten) Machtverhältnissen Auswirkungen auf Machtverhältnisse haben.

Diejenigen, die Technik entwickeln (lassen) – seien es Unternehmen oder der Staat – befinden sich in der Position, Entscheidungen für andere, die diese Technik anwenden bzw. auf die sie angewandt wird, zu treffen: Indem durch Dritte bereitgestellte

[1] ScaDS.AI Dresden/Leipzig, TU Dresden, 01062 Dresden, hermann.diebel-fischer@tu-dresden.de, https://orcid.org/0000-0002-6505-3563
[2] Universität Rostock, Fakultät für Informatik und Elektrotechnik, Institut für Informatik, Didaktik der Informatik, 18051 Rostock, lutz.hellmig@uni-rostock.de

doi:10.18420/inf2022_63

‚Werkzeuge' – seien sie materiell oder immateriell (wie z. B. Ideen) – genutzt werden, unterwerfen sich die Nutzenden fremden Entscheidungen oder Interessen bzw. werden diesen unterworfen. Das passiert ständig und ist nicht *per se* problematisch.

Problematisch kann dies unter mindestens zwei Umständen werden, (1.) wenn die Betroffenen, obwohl sie es wollen, nicht nachvollziehen können, wie die Entscheidung zustande gekommen ist und welche Parameter sie ändern müssen, um ein anderes Ergebnis herbeizuführen und (2.) wenn den Betroffenen (zunächst) nicht deutlich wird, wie durch Technik ein Einfluss auf sie ausgeübt wird. Es mag als ein wenig dramatischer Eingriff aufgefasst werden, wenn eine Software Hinweise dazu einblendet, wie ein Text lesbarer wird (auch wenn damit langfristig der Sprachgebrauch uniformer werden könnte); wenn aber die diffuse Befürchtung eventueller Sanktionen z.B. dafür, dass man sich in elektronischer Kommunikation in einer bestimmten Weise äußert, zur Selbstzensur führt, dann findet eine nicht unerhebliche Einschränkung der eigenen Freiheit statt.

Diese beiden Beispiele befinden sich an verschiedenen Enden eines Spektrums impliziter Technikherrschaft. In ähnlichem Maße, in dem unser Leben einfacher wird, werden dessen Bedingungen komplexer und undurchsichtiger. Obwohl wir heute über wesentlich mehr Informationen verschiedenster Art als alle Generationen vor uns verfügen, sind wir nach wie vor nicht in der Lage, planend zum Wohle aller agieren zu können. Das Gegenteil ist vielmehr der Fall: Wenn Dietmar Dath in klassenkämpferischer Absicht die „Informationsgesellschaft" geißelt, die nichts mehr über „die physikalische noch die ökonomische Wirklichkeit" wisse und sich auch die Technik, die sie nutzt, nicht mehr erklären könne, dann kann man dies als Polemik abtun – oder aber als Aufweis der politischen Dimension von Technik verstehen, die schon lange vor Marx akut war [Dd20, 55].

Dass Entwicklungen auf dem Gebiet der Technik nicht mehr bloß einer Technikfolgenabschätzung bedürfen, zeigt die wachsende Bedeutung der seit dem *Human Genome Project* (1990) etablierten ELSI-Forschung *(ethical, legal, and social implications)*; auch das BMBF-Forschungsprogramm zur Mensch-Technik-Interaktion betont die Bedeutung dieser Forschung als integralen Bestandteil und weist von vornherein – quasi als vorweggenommenes Ergebnis, das zugleich einen normativen Charakter bekommen hat – die Bedeutung menschlicher Kontrolle über Technik aus [Bm18]. Damit wird den impliziten und nur mittelbar wirkenden Machtdimensionen begegnet, deren Folgen auf diesem Weg moderiert werden können. Technik soll dem Menschen nicht fremd sein und auch nicht über ihn hinweg entscheiden, weswegen auch Organisationen wie die IEEE partizipative Konzepte für ihr *Ethically Aligned Design*-Programm verfolgen [IE16].

Diese Bemühungen ergeben nur dann Sinn, wenn die Grundeinsicht lautet, dass Technik weder neutral noch in Gänze beherrschbar ist: Dies gilt insbesondere für Informatiksysteme – und so auch für jenen Bereich, der allgemein KI genannt wird und mittlerweile dahin tendiert, als paradigmatisches Allheilmittel für komplexe Probleme angesehen zu werden.

Auch für die eingangs aufgeführten Beispiele kann behauptet werden, dass bei den Anwender:innen ein Informationsdefizit hinsichtlich dessen, wie sie funktionieren und ihrer gesellschaftlichen Bedeutung besteht. Dieses bedingt vermutlich die Existenz dieser Systeme mit (im positiven wie im negativen Sinne). Vor dem Hintergrund einer als politisch verstandenen Technik, wie Müller-Mall oder Crawford dies für Algorithmen bzw. KI zeigen [Sm20, 37; Kc21], wird schnell der Ruf nach regulierenden Eingriffen laut, die sich selbst der Werkzeuge des Politischen bedienen.

Die Forderung nach einer Ethik für bestimmte Technologien sowie nach Beteiligungsverfahren für die Gestaltung derselben ist jedoch wohlfeil. Damit kann man den der Technik inhärenten Machtstrukturen nur bedingt begegnen, da Verfahren wie die deliberative Demokratie überhaupt keine Garanten dafür sein können, dass am Ende alles gut wird (so auch [Jh22]). Hierfür gibt es zwei Bedingungen: Denn weder ist offensichtlich, in welche materialen Verhältnisse *gut* zu übersetzen ist noch verbürgt die Idee einer – wie auch immer gearteten – Ethik, dass sich über ihre Implementation an verschiedenen Stellen der Relation von konkreten Technologien und Gesellschaft immer schon mehr als eine bloß formale Abarbeitung standardisierter Verfahren bewerkstelligen lässt: Ein als ‚moralisch' qualifiziertes Problem soll bearbeitet oder gar ‚gelöst' werden.

2 Demokratisierung der Technikentwicklung als Scheinlösung

Johannes Himmelreich kritisiert unter dem provokanten Titel *Against "Democratizing AI"* aus politikwissenschaftlicher Perspektive und unter präventiver Abwehr des Vorwurfs, kein Unterstützer der Demokratie zu sein, das Vorhaben, Technikentwicklungen zu demokratisieren. Er identifiziert fünf Gründe, die gegen ein solches Vorhaben sprechen: Zunächst brauche man eine Demokratisierung nur dort, wo es darum gehe, eine Institution zu legitimieren; zweitens würde darüber eine Redundanz erzeugt, da die Regulierung von KI bereits über vorhandene demokratische Strukturen erfolge; drittens sei eine Demokratisierung ressourcenintensiv und bevorzuge jene, deren (Opportunitäts-)kosten für die Teilnahme an einem solchen Prozess gering sind; viertens würden über ein solches Verfahren nur diejenigen beteiligt, die gegenwärtig dazu besonders befähigt sind und die können sich auch irren bzw. durch Interessengruppen in Richtungen getrieben werden, die gerade nicht das Wohl aller im Blick haben; und schließlich fünftens, eine sowohl theoretische als auch praktische Unwirksamkeit einer Demokratisierung mit Blick darauf, diejenigen Missstände abzuwenden, die den Ruf nach einer Demokratisierung allererst erschallen lassen, hervorgerufen dadurch, dass Demokratie an sich „normatively thin" sei [Jh22, 5–11].

Himmelreich macht deutlich, dass sich naheliegende Antworten nicht notwendigerweise als die richtigen erweisen müssen: die Forderung einer Demokratisierung ergibt nur dann Sinn, wenn man in der Anwendung eines demokratischen Verfahrens bereits eine materiale Hinwendung zu dem sieht, was einem selbst oder der eigenen Gruppe als *das*

Gute erscheint. Aber, wie oben bereits erwähnt, folgt aus der Inanspruchnahme von Ethik noch lange nicht, dass es dann für alle *gut* wird.

So votiert auch Müller-Mall nicht für eine simple partizipationsbedingte ‚Lösung' des Problems, dass ein Umgang mit Technik gefunden werden muss, deren politische Bedeutung gravierend ist. Stattdessen spricht sie sich mit Bezug auf das von ihr gewählte Beispiel der Algorithmen für eine Betrachtungsweise aus, die deren Einfluss auf Konstituenten unseres gesellschaftlichen Zusammenlebens, auf „Freiheit und Autonomie" in den Blick nimmt [Sm20, 75f.]. Himmelreich würde dem vermutlich nicht widersprechen, da er fordert, dass für die Demokratisierung von Technik – in seinem Fall: KI – nicht so sehr das *dass*, sondern vielmehr das *wie* in den Blick genommen wird [Jh22]. Aus der Forderung nach einer Ethik für Technik und der Idee, diese über partizipative Verfahren zu realisieren, ergeben sich – auch mit Bezug auf die Dagstuhl-Erklärung [GI16] – Konsequenzen nicht nur für die informatische Bildung, die sowohl für die Lehrenden als auch für die Lernenden ausbuchstabiert werden müssen.

3 Konsequenzen (nicht nur) für die informatische Bildung

Mit Blick auf die GI-Bildungsstandards [GI08; GI16b] kann exemplarisch der Prozessbereich „Modellieren und Implementieren" herangezogen werden. Das Fundament für partizipative Verfahren (nicht nur in der Technikentwicklung) wird in der schulischen Bildung gelegt, indem dort Kompetenzen erworben werden, die die Schüler:innen dazu befähigen, sich konstruktiv zu Fragen der Gesellschaftsgestaltung zu verhalten.

Wenn ich z. B. Ethik algorithmisch abhandle (z. B. sechs Schritte zur Urteilsbildung), dann ist das eine ganz konkrete Konsequenz eines strukturalistisch geprägten Zugangs zu einer Qualität menschlicher Existenz, nämlich Werturteile zu fällen. Diese wird über ihre (Quasi-)Algorithmisierung aber enggeführt. Die Nutzung eines technischen Zugangs verspricht auf den ersten Blick eine über Nachvollziehbarkeit gewonnene Vergleichbarkeit und Verlässlichkeit – diese Standardisierung ist aber nicht nur ein Gewinn, sondern auch ein Preis, den man zahlt: Weil das Modell nicht die Wirklichkeit in Gänze abbildet, besteht die Gefahr, dass bei einer Verselbständigung des Modells bzw. seiner beständigen Anwendung das Modell für die Wirklichkeit gehalten wird. Eine Folge davon besteht darin, dass die Restriktionen, denen das Modell unterliegt, unbeabsichtigt zu realen Restriktionen werden. Für das Beispiel der Ethik bedeutet dies, dass eine komplexe Struktur auf ihre operationalisierbaren Elemente verkürzt wird.

Der Prozessbereich „Modellieren und Implementieren" der GI-Standards [GI16b, 5f.] zielt ab auf das „Abbilden eines Realitätsausschnitts", um dieses Modell als ein Informatiksystem umzusetzen. Durch die Einsicht in die Politizität von Technik und motiviert durch den Inhaltsbereich „Informatik, Mensch und Gesellschaft", könnte dieser Prozessbereich (1.) inhaltlich erweitert und (2.) differenziert werden:

(1.) Die Erweiterung sähe vor, dass schon die Bedingung für eine Modellierung miterfasst werden, d.h. dass gelernt werden soll, zu prüfen, ob sich der „Realitätsausschnitt" grundsätzlich und sinnvoll modellieren lässt und welche möglichen Konsequenzen sich aus dem Modell und für die Realität (und nicht nur für den Ausschnitt) ergeben. Dies würde Schüler:innen dazu befähigen, souveräner mit Modellen und Modellierung umzugehen und bspw. ihre Lebenswelt auf solche Strukturen hin zu untersuchen, verdeckte Modelle und deren Annahmen zu erkennen und ggf. kritisch zu hinterfragen. Dies würde auf einer Meta-Ebene zur Partizipation in der Technikentwicklung beitragen: Je souveräner der Umgang der Bevölkerung mit materieller wie immaterieller Technik ist, desto eher wird eine Fremdbestimmung als eine solche wahrgenommen werden. Dem Monitum Daths, dass wir unsere Existenzbedingungen nur unzureichend reflektieren, könnte damit begegnet werden.

(2.) Für das Modellieren ist zudem ein Schichtenmodell denkbar, das die verschiedenen Abstraktionsebenen in den Blick nimmt, da auch der Realitätsausschnitt selbst schon ein Modell sein kann, dessen Modellhaftigkeit aber unsichtbar geworden ist. So könnte der Realitätsausschnitt nach der Prüfung der Modellierbarkeit differenziert abstrahiert werden: Zunächst geht man bspw. von einem Modell der Gesellschaft aus, aufgrund dessen dann ein Modell eines Sachverhalts gewonnen wird, das dann in ein informatisches Modell übersetzt wird.

Ein Beispiel illustriere dieses Schichtenmodell: (Noch) wird in vielen Datenbanken bzw. Softwareprojekten das Attribut ‚Geschlecht' binär codiert – als Grundlage hierfür dienen tradierte Normvorstellungen von physiologischen Attributen des menschlichen Körpers. Mit einer erweiterten und differenzierten Sicht auf den Modellbildungsprozess kann zunächst die Frage gestellt werden, ob die Erfassung dieses Attributs aus ethischen Erwägungen überhaupt opportun ist. Dies vorausgesetzt folgt die Erörterung der sozial und biologisch geprägten Semantik des Attributs ‚Geschlecht'. In welcher Form soll die binäre Sicht auf das Attribut ‚Geschlecht' überwunden werden, d.h. welche Abstraktionen werden zunächst auf der Sachebene vorgenommen: Als dreiwertige (m/w/d) Auswahlliste oder als Bezeichnung frei wählbar? Die Entscheidung hierfür ist sowohl gesellschaftlich als auch technisch indiziert, schließlich hat die Festlegung des repräsentierenden Datentyps im informatischen Modell auch Auswirkungen auf die Möglichkeiten der automatischen Datenverarbeitung und ihrer Auswertung und damit auch wieder auf die Interpretation und Wirkung der Daten in der Gesellschaft. Der informatische Modellierungsprozess kann nicht losgelöst vom sinngebenden Sachverhalt erfolgen. Handelt es sich um konkrete Sachverhalte aus dem Alltagsumfeld der Schüler:innen, ist dieser nicht nur von politischen und ethischen Fragestellungen und Modellierungen abhängig, sondern beeinflusst diese wiederum im Weiteren.

Die Form eines Schichtenmodells zur Beschreibung dieser Abhängigkeiten wurde gewählt, um das Primat der gesellschaftlichen Verhältnisse gegenüber der technischen Realisierung klarzustellen. Nichtsdestotrotz ist das Verhältnis zwischen diesen Schichten von komplexen Interdependenzen geprägt.

So wie technische Bildung immer die politische und ethische Dimension mitbedenken muss, so kann man heute politische und ethische Bildung nicht ohne technische Bildung unterrichten. Dies betrifft nicht nur die technischen Bedingungen partizipativer Verfahren, sondern setzt schon dort ein, wo die technische Überformung grundlegender Ideen der Gesellschaftsgestaltung unsichtbar geworden ist – z. B. dann, wenn Bildungsprozesse anhand technischer Restriktionen strukturiert werden („Das kann das System leider nicht abbilden.“) oder die Aufgabe der Ethik als ein algorithmisierbares Problemlösen missverstanden wird. Ein souveräner Umgang mit den Modellen und der Modellierung kann durch eine enge Verzahnung technischer (informatischer) Bildung mit politischer und ethischer Bildung erreicht werden. Dadurch werden unsichtbare Verhältnisse für sehr viel mehr Menschen sichtbar, was den Ruf nach einer Ethik für Technik nicht mit Regulierung, sondern mit Kompetenzen beantwortet.

4 Danksagung

Der Anteil von Hermann Diebel-Fischer an dieser Arbeit wurde unterstützt durch das Bundesministerium für Bildung und Forschung (BMBF, 01IS18026A-D) im Rahmen der Finanzierung des Kompetenzzentrums ScaDS.AI Dresden/Leipzig.

5 Literaturverzeichnis

[Bm18] Bundesministerium für Bildung und Forschung: Technik zum Menschen bringen. Forschungsprogramm zur Mensch-Technik-Interaktion. Berlin, 2018.

[Dd20] Dath, Dietmar: Woher kommt das Neue? Jacobin, 2/2020, 52–65, 2020.

[GI08] Gesellschaft für Informatik: Grundsätze und Standards für die Informatik in der Schule. Bildungsstandards Informatik für die Sekundarstufe I, 2008.

[GI16a] Gesellschaft für Informatik: Dagstuhl-Erklärung. Bildung in der digitalen vernetzten Welt. 2016.

[GI16b] Gesellschaft für Informatik: Bildungsstandards Informatik für die Sekundarstufe II, 2016.

[IE16] IEEE: The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems. Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems. 2016

[Jh22] Himmelreich, Johannes: Against “Democratizing AI”. AI & Society, 2022.

[Kc21] Kate Crawford: Atlas of AI. Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press, New Haven, 2021.

[Sm20] Sabine Müller-Mall: Freiheit und Kalkül. Die Politik der Algorithmen. Reclam, Ditzingen, 2020.

12. Betriebliche Umweltinformationssysteme (BUIS-Tage 2022)

A New Approach to Determine Cycling KPIs in the Context of Behavioral Data

Harish Moturu (harish.moturu@uni-oldenburg.de),

Jorge Marx Gómez (jorge.marx.gomez@uni-oldenburg.de),

Johannes Schering (johannes.schering@uni-oldenburg.de)[1]

Abstract: The Corona pandemic has brought a great change in the mobility sector especially in European countries. As a consequence many cities invested much more budget in bike infrastructure than before while the percentage of people using bicycle as the mode of transport for every day routine has been rapidly increased at many places . Available data on cycling use, traffic safety etc. can support the decision making process. We dont't know much about distraction and stress factors when using the bike in a city because such kind of data was not available until now. In the beginning of this paper we present the current state-of-the-art in field of cycling KPIs. What is new in the cycling research domain are cycling KPIs that consider behavioral data. A set of new KPIs in this specific field are introduced in this contribution. As part of the SmartHelm project EEG, eyetracking and GPS data of cargo bike riders was collected and processed. The paper describes the preliminary results obtained from implementing the methodology to calculate behaviour based new cycling KPIs. The conclusion states how the KPIs can be applied in various fields and how it can be adjusted and expanded in the future to the specific stakeholder requirements.

Keywords: KPI, Behavioral Data, Smart Cycling, Cargo Bike Logistics

1 Introduction

Many cities have discovered the potential of cycling promotion to create a liveable and sustainable city [SF21], [IO18]. More interest in cycling leads to the question how bike infrastructure can be made more convenient and adjusted to the user demands. To support this process, cycling data driven initiatives were initiated in the 2010s [Sc19]. In principle, according to the report published by European Commission on Indicators for Sustainable Cities [Co18], [De22] Key Performance Indicators (KPIs) can be defined as Urban sustainability indicators, which are the tools that allow city planners, city managers and policymakers to gauge the socio-economic and environmental impact. The related KPIs [Fi90] can help to learn more about the cycling behavior, traffic safety or the perceived cycling attractiveness in a city. As an example, the calculated numbers can show us accident hotspots or the amount of bike trips in a municipal area. What we have no information about are distraction factors when using the bike in a city. In a digitized and motorized world distractions in the traffic situation are very common. For instance, mobile phone use plays an

[1] University of Oldenburg, Business Informatics, Straße, 26129 Oldenburg, Germany

doi:10.18420/inf2022_64

increasing role in cycling accidents [De10]. New cycling KPIs that specify on biophysical data could provide new insights in traffic safety situations. The SmartHelm[2] project tries to make available EEG and eyetracking sensor data in the specific usecase of bicycle logistics. As part of this work new biophysical cycling KPIs were created. These numbers that are presented in this contribution should improve the understanding of distraction factors in the biking scenario. The paper will be structured in the following sections. The first part of this contribution is a literature research in the field of existing cycling KPIs. As we will see, there are already quite a lot of bicycle specific indicators available. At this moment these are not related to distractions and behavioral data. The following chapter describes the available data sources that are the basis for the creation of behavioral cycling data KPIs. We will describe the challenges in the data processing process. The joining of heterogeneous data streams is one of the major challenges. The following chapter presents the preliminary KPIs and how these are calculated. In the outlook we will discuss what will be the next working steps. Based on the feedback of domain experts the KPIs will be further developed and adjusted to make a contribution to the improvement of cycling conditions.

2 State of the Art

The state of the art in the cycling data domain includes different approaches to create KPIs to make these related numbers available, comparable and understandable. Open available bike sharing systems can be a good starting point. The paper from [Ya17] focuses on the impact caused by public and private funded bike sharing models on sustainability index, it proposes some new metrics with respect to cycling to promote the sustainable mobility. These performance metrics are calculated based upon the data collected from the bike sharing applications. One of the significant metric proposed in this work [Ya17] is Overall Bike Effectiveness (OBE), which is derived from the Overall Environment Effectiveness (OEE). The fundamental ground factors taken into consideration to calculate OEE are availability, performance and quality of the transport modes in Bike Sharing System (BSS). In addition other different metrics such as Overall Transportation Effectiveness (OTE), Overall Vehicle Effectiveness (OVE), and Overall Labor Effectiveness (OLE) are also elaborated depending upon the various criteria which come under urban mobility and BSS. Although this paper does not explicitly calculate performance indicators in the domain from human computer interaction, but it made an attempt to present few performance indicators in cycling mobility. The next paper in the literature [As20] discusses about the Context-Aware Human Activity Recognition (CAHAR) method. In this paper the main focus is to compare performance of various conventional machine learning algorithms on Human Activity Recognition (HAR), which means human behavioural performance indicators are measured during different activities from sensor data. Accelerometer data collected from smartphone sensor is preprocessed to evaluate various features in the data. The feature extraction process involves in finding out different time domain features which will indeed

[2] https://smart-helm.com/

result to recognize activities from the processed data. Few of the primary activities and their corresponding behavioral contexts recognized from the accelerometer data are as follows. In the primary activity series first activity refers to lying down from which the behavioural context is classified as either of these three sleeping, surfing in the internet or watching TV. The second activity is standing which corresponds to the indoor, outdoor, shopping, and talking are the behavioral contexts. The next primary activity in the list is bicycling, here the behavior contexts recognised is exercise. Similarly running as well as standing are classified as exercise and indoor respectively. In an overall some of the basic indicators resulted from the study [As20] are common machine learning algorithm evaluation metrics such as precision, sensitivity, specificity and accuracy. Although it does not introduce new performance indicators in this field, but it deals with human interaction with mobility and summarizes the some of the performance indicators in context to different human activities scenarios.

In general, simultaneously recognizing human behavior and understanding an users mobility from sensor data is a complex process, where the following paper [Zh08] introduced new approach to understand mobility behavior of users based on GPS Data. In research study [Zh08] an in depth mobility as well as behavior analysis is carried out on 65 participants who rode bicycles for a period of 10 months in the city of Shanghai, China. Here, the mobility data is acquired with GPS sensor device and change point-based graph approach is applied to recognize the user activity, where a point-based graph classifies the activities based upon the user head movement into walking, driving, using public transport etc.,. For classification typical machine learning algorithms are bought into application, where pre-processing follows with feature extraction by application of various algorithms. The performance indicators resulted from this paper are Stoping Rate (SR), Heading Change Rate (HCR), and Velocity Change Rate (VCR) etc. these metrics have been calculated in terms of behavioral context. The three metrics SR, HCR, and VCR have a relation with user behavior and mobility. Multimer technology described in the paper [DKG18] is perhaps one of the first spatial technology that collects biophysical data, like brainwave and heart rate data, and analyzes it with both spatiotemporal and neurophysiological tools, the report and results are presented in Multimer [DKG18] study conducted on 101 subjects and the data recorded from these subjects. Based upon the the data collected from multimer technology various validation results are presented such as 'Mediatation factor', 'Attention factor' and precalculated band powers. The results are categorised depending on Pedestrians (Peds), Bicycle Couriers (Cour), and Bicycle Commuters (Comm). The following papers [SM22], [Sc20] exclusively describes about the KPIs calculated from the bicycle data[3] where the data is acquired in different contexts such as counting data, parking data, snifferbike and bicycling mobile application. Some of the the important outcomes from these papers are the average distances and durations of cycling trips. Bicycle trips measured from smartphone users tend to be longer than that of collected with mobile sensors [SM22]. In addition, here weather based KPIs are also calculated which showed bad weather conditions reflect to short bike trips with less effect of speed.

[3] https://www.bicycle-data.de/city-analysis/

3 Methodology

Firstly data collected from the experiments in the SmartHelm project is processed and with standard Extract, Transfer, and Load (ETL) methodology is stored in a data management system published in a seperate paper [MSM22]. The various types of data acquired from the experiments is electroencephalography (EEG) data, Eye-tracking data, attention information, speech and GPS information. [MSM22] publication describes more elaborately about the data stored in the data management system and external data stored in the system. In principle the complete data in data management system is joined on the time domain basis to calculate various performance indicators.

Now describing about the individual challenges and respective solutions while deriving the KPIs in the context from the cycling and behavioral data. Primarily challenges in handling the data i.e frequency difference which means the sampling rate of each data to make combined queries. In principle data acquired from multi-modal systems contains relatively more heterogeneity in comparison to other data sources, for the purpose of data analysis and evaluation a fine structured and constituted data source should be the fundamental requirement. As explained in the beginning of this chapter about the comprehensiveness in the data sources, which means contextually one data type differs completely from another. The main reason behind this heterogeneity is due to the type of the data acquired from the multi modal system. In the SmartHelm project the main focus is to analyze the user attention during the experiment. The test activity SmartHelm Prototype are described in this paper [Ch20]. The SmartHelm prototype records data from various data streams as mentioned from the multi-modal set-up. Now here the challenge for data analysis and data visualization is connecting the data streams acquired under different frequency in order to find out the behavioural characteristics of the participant with respect to the geographical location. For performing spatial and temporal analysis all the data sources needed to be concatenated with each other. As Individual data streams recorded from SmartHelm prototype have different sampling rate, a fact data creation concept is implemented to overcome this challenge. The name fact data determines one big table constructed by combining the significant data sources to implement data analysis.

The principle methodology involved in formulating the fact data for joining individual data segments with one another is batch processing, which means fact data for each experiment is constructed and then loaded into the database as batches. During this batch processing data with lowest sampling frequency is taken as base and joined with adjacent data with the greater sampling frequency which means in the current data case, location data has the lowest sampling frequency and it's adjacent data with next greater sampling frequency is eye-tracking data. Therefore a new column is created from timestamps from the base location data where each successive timestamp value will be created on the new adjacent column. Now all the values in the reference data i.e. eye tracking which fall in the range of the old timestamps column and new timestamps column are joined with the location. Further exactly same kind of merging is followed now by considering combined eye tracking, GPS data as base structure and EEG as reference structure.

4 Findings

Measurement of KPIs for Behavioral data. The KPIs defined below are the preliminary findings (currently work in progress) obtained from querying the SmartHelm data which consists both bio-physiological and mobility factors. The KPIs described below are very new kind in the cycling KPIs domain, because in general either complete mobility related KPIs are calculated or sustainabilty performance indicators are analyzed but in this paper the first results are presented to study the user distraction related factors during the motion.

1. **Mean EEG Channel-MEC:** In total 7 channels of EEG signal values are recorded as part of each experiment with a frequency of 250 Hz. The performance indicator MEC calculates the average electrode signal from each channel for the duration of one complete session. Here session can be described as the time duration for one participant to participate in a single experiment, usually a session lasts in between 10 to 12 minutes. This performance indicator signifies an approximate range of EEG values obtained from the experiment. The scientific significance of each EEG signal value cannot be interpreted with the MEC performance indicator, it is just only an index to know the mean channel values. In addition to this performance indicator some basic factors such as maximum and minimum values of the EEG channels are also calculated. 1 shows an example KPI from calculated for one session.

Participant ID	**Mean EEG channel 1**	**Mean EEG channel 2**	**Mean EEG channel 3**	**Mean EEG channel 4**
B (Anonym)	0.0036 microvolt	-0.022 microvolt	-0.0214 microvolt	0.036 microvolt
	Mean EEG channel 5	**Mean EEG channel 6**	**Mean EEG channel 7**	**Experiment location**
	-0.0884 microvolt	0.031 microvolt	-0.044 microvolt	Bremen

Tab. 1: Mean EEG Channel value

2. **Mean Object Distance (MOD):** is an index measured from the recordings received from the Augmented Reality Glass (HoloLens2) that is integrated in the SmartHelm. Actually here (MOD) is calculated from the eye gaze direction coordinates values obtained from the Hololens2 in x, y, and z axis as well eye gaze origin coordinates along x, y, and z axis. MOD helps in finding out the three dimensional mean distance of a object viewed from the point of contact. For example, during the experiment the participant rides a bike and focuses on a particular object, MOD calculates mean distance from the participant view to the object. 2 shows an example KPI from calculated for one session.

Participant_id	**Mean object distance along x axis**	**Mean object distance along y axis**	**Mean object distance along z axis**	**Mean 3D object distance from the origin**
B (Anonym)	27.5 cm	33.3 cm	63.4 cm	67.07 cm

Tab. 2: Mean Object Distance

3. **Biking Altitude (BA):** The performance indicator Biking Altitude (BA) 3 compares the highest and least biking altitude between two experiment locations. As of now this

KPI just only calculates the mean biking altitude of the participant but further more it will be connected with additional factors to know the behavioral context relating to the average elevation at which the experiments are conducted.

Participant_id	Experiment Location	Maximium Biking Altitude	Minimium Biking Altitude
A (Anonym)	Bremen	72 m	49 m
B (Anonym)	Oldenburg	90 m	54 m

Tab. 3: Biking Altitude Comparision

4. The **Variable Velocity with Altitude (VVA):** The last KPI 4 in the order signifies the change in the biking speed with respect to the altitude of the location. From the 4 it can be clearly observed in the location 1 biking speed is less than then the biking speed in location 2 where the mean altitude in location 1 is greater than the location, which specifies that biking speed is dependent with attitude. As mentioned earlier the future work will consider an additional factor that is attention level so that the KPI can be extended in behavior context.

Participant_id	Experiment Location	Mean Altitude	Mean Biking speed
A (Anonym)	Oldenburg	60.5 m	9.6 Kmh
B (Anonym)	Bremen	51. m	12.4 kmh

Tab. 4: Variable Mean Velocity versus Altitude

5 Conclusion and Outlook

The KPIs discussed in this paper are the preliminary findings from the data obtained from SmartHelm prototype 1. These findings found to be very intuitive in understanding the principal metrics from the inter disciplinary data i.e the data recorded in the context of both behavioural and mobility aspect. However, an explicit KPI defining the distraction of the bicycle rider is not yet mentioned, because the values specifying the annotated distractions are not integrated in the data management system. After the integration of annotated distractions, then KPIs with distractions will be calculated and published for the future publication. In this paper we presented some of the results from an on going work. Some of the future planned KPIs are as follows. The first step to integrate the attention data points with the existing data and then calculate few relevant KPIs (e.g. distraction location, timestamp and distraction counts per route) accordingly. The next set of the KPIs will take into consideration of the weather data as well to determine the influence of rain, wind and sun shine etc., with respective to the cycling behavior. Finally in the last phase after the data received from the SmartHelm prototype 2 speech related KPIs are also in planning to calculate. At this development stage the relevance of the newly created KPIs on awareness sensitive cycling data is not clear. The KPIs will be adjusted to the stakeholders (bicycle logistics domain, traffic planners, and experts in cognitive systems) requirements

in the future. On the one hand, the results should contribute to improve knowledge about distraction and stress factors in the cargo bike logistics domain. The biophysical cycling KPIs will be evaluated on its benefit for the bicycle delivery service. On the other hand, the expertise on experts in the field of Cognitive Systems will be applied. These will bring their perspective to the processing of EEG and eyetracking data.

References

[As20] Asim, Y.; Azam, M. A.; Ehatisham-ul-Haq, M.; Naeem, U.; Khalid, A.: Context-aware human activity recognition (CAHAR) in-the-Wild using smartphone accelerometer. IEEE Sensors Journal 20/8, pp. 4361–4371, 2020.

[Ch20] Chatterjee, S.; Scheck, K.; Küster, D.; Putze, F.; Moturu, H.; Schering, J.; Marx Gómez, J.; Schultz, T.: SmartHelm: Towards Multimodal Detection of Attention in an Outdoor Augmented Reality Biking Scenario./, pp. 426–432, 2020.

[Co18] Commission, R. E.: IN-DEPTH REPORT: Indicators for Sustainable Cities. Science for Environment Policy 12/1, pp. 5–6, 2018, URL: https://bit.ly/3aMjgMK.

[De10] De Waard, D.; Schepers, P.; Ormel, W.; Brookhuis, K.: Mobile phone use while cycling: Incidence and effects on behaviour and safety. Ergonomics 53/1, pp. 30–42, 2010.

[De22] De Sanctis, M.; Iovino, L.; Rossi, M. T.; Wimmer, M.: MIKADO: a smart city KPIs assessment modeling framework. Software and Systems Modeling 21/1, pp. 281–309, 2022.

[DKG18] Ducao, A.; Koen, I.; Guo, Z.: Multimer: validating multimodal, cognitive data in the city: towards a model of how the urban environment influences streetscape users. In: Proceedings of the Workshop on Modeling Cognitive Processes from Multimodal Data. Pp. 1–8, 2018.

[Fi90] Fitz-Gibbon, C. T.: Performance indicators. Multilingual Matters, 1990.

[IO18] Ibsen, M. E.; Olesen, K.: Bicycle urbanism as a competitive advantage in the neoliberal age: the case of bicycle promotion in Portland. International planning studies 23/2, pp. 210–224, 2018.

[MSM22] Moturu, H.; Schering, J.; Marx Gómez, J.: SmartHelm: Design and Implementation of Open Source based Behavioral Data Repository. In: The Twelfth International Conference on Mobile Services, Resources, and Users MOBILITY 2022. In Proceedings, 2022.

[Sc19] Schering, J.: Vision 2045: Potentiale des Radverkehrs im Transformationsprozess der autogerechten Stadt zur nachhaltig orientierten Smart City. In: Smart Cities/Smart Regions–Technische, wirtschaftliche und gesellschaftliche Innovationen. Springer, pp. 211–225, 2019.

[Sc20] Schering, J.; Janßen, C.; Dmitriyev, V.; Marx Gómez, J.; Stehno, C.; Pelzner, K.; Bankowsky, R.; Hentschel, R.: ECOSense and its preliminary findings: Collection and analysis of bicycle sensor data. In: Kamilaris, A.; Wohlgemuth, V.; Karatzas, K.; Athanasiadis, I. (ed.). Environmental Informatics – New perspectives in Environmental Information Systems: Transport, Sensors, Recycling. Proceedings of the 34th EnviroInfo conference 34/, pp. 145–153, 2020.

[SF21] Scorza, F.; Fortunato, G.: Cyclable cities: building feasible scenario through urban space morphology assessment. Journal of Urban Planning and Development 147/4, p. 05021039, 2021.

[SM22] Schering, J.; Marx Gómez, J.: The BITS Project–Making cycling data available and comparable on a European scale. Transportation Research Procedia 60/, pp. 424–431, 2022.

[Ya17] Yahya, B. N.: Overall bike effectiveness as a sustainability metric for bike sharing systems. Sustainability 9/11, p. 2070, 2017.

[Zh08] Zheng, Y.; Li, Q.; Chen, Y.; Xie, X.; Ma, W.-Y.: Understanding mobility based on GPS data. In: Proceedings of the 10th international conference on Ubiquitous computing. Pp. 312–321, 2008.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 771

Elektromobilität in Unternehmen - Schwerpunkte und Methoden der Forschung

Bettina Steden[1], Jorge Marx Gómez[2]

Abstract: Die Relevanz alternativer Antriebe steigt. Auch für Unternehmen wird eine Abkehr von konventionellen Antrieben zunehmend interessant. Die bekannteste Alternative zu konventionellen Fahrzeugen stellt Elektromobilität dar. Die Möglichkeit, Elektromobilität in verschiedenen Unternehmen einzusetzen wurde folglich auch in der Forschung vielfach betrachtet. In diesem Paper wird daher mit Hilfe einer Literaturanalyse untersucht, wo die Forschungsschwerpunkte bei dem Einsatz von Elektromobilität liegen. Es kann gezeigt werden, dass fünf Themenschwerpunkte existieren. Dies sind zum einen Entscheidungsunterstützung und die Auswirkungen einer Elektrifizierung. Zum anderen werden die Akzeptanz von Elektromobilität, das Laden sowie das Potential, das Elektromobilität bietet, schwerpunktmäßig untersucht.

Keywords: Elektromobilität, Unternehmen, nachhaltige Mobilität

1 Einleitung

Elektromobilität gewinnt zunehmend an Bedeutung in Politik und Gesellschaft. So stiegen die Neuzulassungen von Elektrofahrzeugen von 3 % im Januar 2020 auf 14 % im Dezember 2020 [Di]. Auch in Unternehmen steigt das Interesse, Fahrzeugflotten nachhaltiger zu gestalten. Die beschränkte Reichweite und die längere Ladedauer im Vergleich zu herkömmlichen Fahrzeugen erschwert jedoch den Umstieg auf Elektromobilität. Die Einsatzmöglichkeiten in verschiedenen Unternehmen wurden folglich auch in der Forschung vielfach betrachtet. Die untersuchten Unternehmen stammen dabei aus den verschiedensten Sektoren wie zum Beispiel dem städtischen Lieferverkehr oder dem öffentlichen Sektor.
In diesem Paper wird daher mittels einer Literaturanalyse dargestellt, wo die Forschungsschwerpunkte bei dem Einsatz von Elektromobilität liegen. Es sollen die folgenden Fragestellungen beantwortet werden: Welche Problemstellungen werden untersucht? Welche Methoden kommen schwerpunktmäßig zum Einsatz? Ein besonderer Fokus soll auf Veröffentlichungen gelegt werden, die Elektromobilität im praktischen Einsatz untersuchen. Auf diese Weise können Rückschlüsse auf die Herausforderungen bei der Einführung von Elektromobilität in Unternehmen gezogen werden.

[1] Universität Oldenburg, VLBA, Ammerländer Heerstraße 114-118, 26129 Oldenburg, Deutschland bettina.steden@uni-oldenburg.de
[2] Universität Oldenburg, VLBA, Ammerländer Heerstraße 114-118, 26129 Oldenburg, Deutschland jorge.marx.gomez@uni-oldenburg.de

doi:10.18420/inf2022_65

1.1 Verwandte Arbeiten

In [GHR21] werden Forschungsergebnisse der Jahre 2016-2020 in dem Bereich der unternehmensbezogenen Mobilität betrachtet. Sie können feststellen, dass Forschungsschwerpunkte auf verkehrsplanerischen Ansätzen und politischen Anreizen sowie auf den Stakeholdern liegen. Die in [GHR21] untersuchte Literatur beschränkt sich jedoch nicht auf Elektromobilität. [Pa21] führen ein Literaturreview im Bereich der Last-Mile-Logistik durch, um die wesentlichen Fragestellungen und Herausforderungen bei der Einführung von Elektrofahrzeugen zu ermitteln. Sie identifizieren die drei Forschungsschwerpunkte Nachhaltigkeit, Politik und Optimierungsprobleme. In [Wa18] wird ebenfalls der Fokus der Literatur bei Elektromobilität im städtischen Lieferverkehr untersucht. Es zeigt sich, dass der Großteil der untersuchten Veröffentlichungen die Realisierbarkeit von Elektrofahrzeugen betrachtet, die Unterstützung von Stakeholdern hingegen deutlich seltener adressiert wird. [Ju16] führen ein Literaturreview durch, um die ökologischen sowie planerischen und operativen Herausforderungen bei der Nutzung von Elektromobilität zu identifizieren. Ein besonderer Fokus liegt hier auf Vehicle Routing Problemen.

2 Methodik

Zur Beantwortung der Fragestellungen wurde ein systematisches Literaturreview nach [Br15] durchgeführt. Die Datenbanken Scopus, Web of Science, Science Direct und IEEE wurden mit dem Suchbegriff „Elektromobilität“ in Verbindung mit „Unternehmen“, sowie Synonymen dieses Begriffs, durchsucht. Die Suchbegriffe und Ergebnisse können Tabelle 1 entnommen werden. Die in dieser Tabelle dargestellten Ergebnisse sind bereits gefiltert. Publikationen, die dem Titel nach nicht im Zusammenhang mit dem Thema „Elektromobilität in Unternehmen“ stehen, wurden zuvor aussortiert. So enthielten die Ergebnisse viele medizinische Publikationen, die ausgeschlossen werden konnten. Bezüglich der Erscheinungsjahre gibt es keine Einschränkung.
Von den Ergebnissen sind insgesamt 25 Paper mit in die weitere Analyse eingeflossen. Über eine Vorwärts- und Rückwärtssuche mit diesen Veröffentlichungen konnten 31 weitere relevante Publikationen identifiziert werden. Nach einer Analyse der Paper konnten induktiv Kategorien zur Einteilung dieser in Forschungsschwerpunkte gebildet werden.

Electromobility AND	Company	Car Fleet	Corporation	Business Enterprise	Business	Commercial
Scopus	11	9	0	1	4	8
Web of Science	2	4	0	1	2	5
Science Direct	3	6	1	2	0	13
IEEE Xplore	2	2	0	0	0	2

Tab. 1: Ergebnisse im Zusammenhang mit den Suchbegriffen nach Ausschluss irrelevanter Themen

2.1 Limitationen

Es gibt weit mehr Themenbereiche, die sich mit der Nutzung von Elektromobilität beschäftigen, als hier abgebildet werden können. Auf Grund der beschränkten Reichweite von

Elektrofahrzeugen bieten sich Vehicle Routing Probleme an. Da diese jedoch genug Umfang für eine eigenständige Literaturanalyse bilden, wurden Publikationen, die allein dieses Thema betrachten, nicht mit aufgenommen. Für ein Review in diesem Bereich siehe zum Beispiel [KDC21]. Ebenso kann hier nicht die Bandbreite der Veröffentlichungen abgedeckt werden, die sich allein mit Kostenvorteilen oder Unterschieden in den Treibhausgasemissionen beschäftigen. Bei der Auswahl der Paper wurden einzig Publikationen, die Carsharing Unternehmen betrachten, nicht weiter berücksichtigt, da diese Unternehmen nochmals eine besondere Struktur bieten. Davon abgesehen fand im Vorfeld keine Einschränkung auf bestimmte Unternehmen statt.

3 Ergebnisse

Veröffentlichung	Ziel
Entscheidungsunterstützung (9)	
[WBL19]	Ganzheitliche Lösung zur Nutzung von Elektromobilität
[Kr21]	Hierarchie zur Elektrifizierung von Fahrzeugen
[HE20], [Wą17], [KIM19]	Multikriterielle Entscheidungsanalyse
[Te22] , [Ma22], [IL19]	Optimierung Fahrzeugpoolzusammenstellung
[Be15], [HE20], [KIM19], [Wą17]	Unterstützung des Planungsprozess
Potentiale und Herausforderungen (22)	
[Iw21]	Effizienz elektrischer Transportfahrzeuge
[Iw19], [An21], [QC21], [Vu21], [Di21], [Ga21]	Hindernisse und Herausforderungen Nutzung Elektrofahrzeuge (EVs)
[Qu16], [FW15], [Jo20], [Wą17]	Machbarkeit Nutzung EVs
[TKT16]	Operative Eignung EVs
[BL16], [FW15], [Iw19], [Fi18], [Be17], [KKR13], [Ga21], [WHA14], [Ro22], [Gn13]	Potential der Integration von EVs
[Ch17], [KRZ16]	Potential verschiedener Sektoren
[Ma16]	Technische Merkmale mit Einfluss auf Effizienz von EVs
Motivation und Akzeptanz (15)	
[GDS18], [Vu21]	Akzeptanz von EVs
[Qu16], [KH14], [Gi19]	Einstellung gegenüber der Nutzung von Elektromobilität
[GDW18], [İÇK21], [Ka16], [Bo18]	Faktoren mit Einfluss auf Inanspruchnahme/Akzeptanz von EVs in Fahrzeugflotten
[SSKP21], [Vu21], [Si14]	Faktoren mit Auswirkung auf Interesse an alternativen Antrieben
[KH14]	Identifikation von Early Adoptern
[CK12]	Informationen auf Grundlage derer Kaufentscheidungen von Firmen getroffen werden
[Ja21], [Vu21], [LMV16], [Si14]	Motivation Einführung EVs
[Mi19]	Projekte, um Nutzung Elektromobilität voranzubringen
Auswirkungen der Elektrifizierung (12)	
[MBC14], [Iw21]	Auswirkung Nutzung von EVs auf die Energieeffizienz
[MBC14]	Auswirkung Nutzung von EVs auf den Verkehr
[Ts20], [Ar20]	Kostenvorteile EVs
[Gh21]	Minimierung Total cost of Ownership
[Be15], [MBC14], [KKR13], [An20], [Bo18], [PP20], [BL16], [Ma18]	Ökonomische und ökologische Auswirkungen
Laden (8)	
[FFR20], [WBL19]	Analyse Ladeverhalten
[Ar20]	Effekte einer CO_2 Bepreisung
[Ro22]	Einfluss Ladeinfrastruktur auf Elektrifizierungspotential
[KHQ21], [Mä21]	Ladestrategien, Simulation von Ladevorgängen
[Bo18]	Optimierung elektrischer Infrastruktur
[Te22]	Vehicle to Grid Erlöse
Sonstiges (3)	
[Os14], [KO16]	Carsharing zwischen verschiedenen Unternehmen
[Na21]	Konzipierung und Umsetzung elektrisches Lieferfahrzeug

Tab. 2: Ergebnisse der Literaturanalyse

Die Ergebnisse stammen aus den Jahren 2012-2022, wobei ein Großteil seit 2019 erschienen ist, siehe Abbildung 2. Eine Einteilung der Ergebnisse nach Sektoren ist in Abbildung 1 dargestellt. Es wird insbesondere der städtische Lieferverkehr untersucht, da er sich durch planbare Strecken und kurze Wege auszeichnet. Eine ebenso große Rolle spielen allgemeine Firmenflotten.

Eine zusammenfassende Darstellung der Paper und deren Ziele ist in Tabelle 2 dargestellt. Die Ergebnisse können hauptsächlich den Themenbereichen *Entscheidungsunterstützung*, *Auswirkungen der Elektrifizierung*, *Motivation und Akzeptanz*, *Laden* sowie *Potentiale und Herausforderungen* zugeordnet werden. In den nachfolgenden Abschnitten werden die Themenbereiche anhand ausgewählter Publikationen näher erläutert. Es werden diejenigen Veröffentlichungen gewählt, die über Piloteinführungen oder reale Flottendaten das Vorgehen evaluieren. Dies ermöglicht eine Ableitung von Herausforderungen an eine Einführung von Elektromobilität.

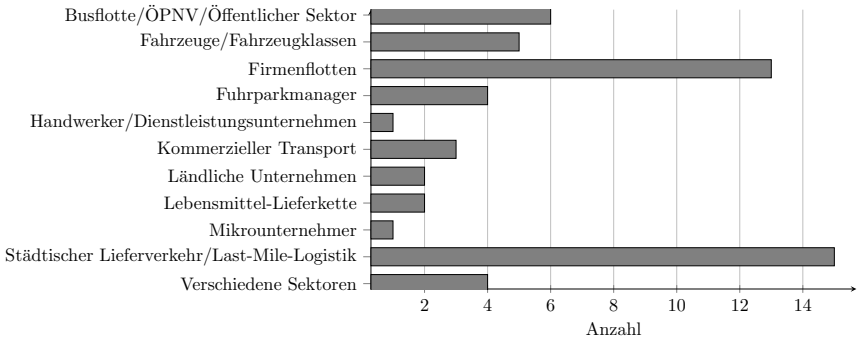


Abb. 1: Veröffentlichungen nach Sektoren (eigene Darstellung)

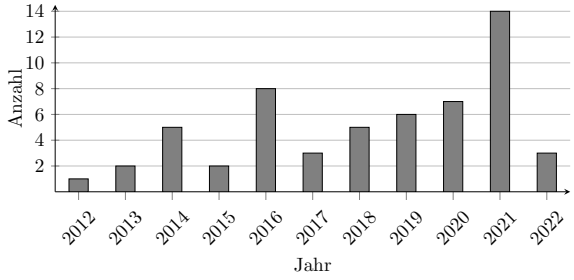


Abb. 2: Veröffentlichungen nach Erscheinungsjahren (eigene Darstellung)

Entscheidungsunterstützung. Der Themenbereich Entscheidungsunterstützung umfasst die Unterstützung des Planungsprozesses bei der Einführung von Elektromobilität sowie die Ermittlung der optimalen Poolzusammenstellung. Hierzu werden in erster Linie mathematische Modelle genutzt. Eine wichtige Rolle spielt die multikriterielle Entscheidungsunterstützung, da meist viele zum Teil konkurrierende Ziele und Kriterien betrachtet werden. So nutzen [Wą17, KIM19] multikriterielle Entscheidungsanalyse, um den Einsatz ausgewählter Fahrzeuge im Lieferverkehr zu bewerten. Die dort definierten Bewertungskriterien lassen sich in die Kategorien Leistung, Batterie, Motor und Anschaffungspreis einteilen. Zur Bewertung der Fahrzeuge anhand dieser Kriterien, werden Promethee II [BVM86] und TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) [PP18] genutzt. In Feldtests wurden [KIM19] mit dem Problem konfrontiert, dass nicht

alle Daten zu den aufgestellten Kategorien zur Verfügung stehen und demzufolge mit Durchschnittswerten gearbeitet werden muss. [HE20] nutzen zur Evaluation des Einsatzes von Elektromobilität in Busflotten ebenfalls TOPSIS. Bei der Anwendung des Ansatzes auf Busflotten verschiedener Metropolen stellen sie fest, dass vielfach keine klare Dominanz einer Handlungsalternative vorliegt.
Um die optimale Flottenzusammenstellung zu ermitteln, wenden [Ma22] auf zwei Flotten das stochastische Fleet-Size-and-Mix Problem an. Hierbei zeigen sich unterschiedliche Nutzungszeiten für Sommer und Winter, die bei der Planung berücksichtigt werden müssen. [KKR13] entwickeln ein Simulationsmodell anhand dessen das Potential für die Nutzung von Elektromobilität in Unternehmensflotten ermittelt werden kann. Das Potential wird mittels zurückgelegter Strecken und der Zeit zwischen zwei Buchungen auf einem Fahrzeug bestimmt. Die praktische Anwendbarkeit dieses Ansatzes zeigt sich in der Umsetzung als Geschäftsmodell [Fr]. [WBL19] entwickeln in Form eines mehrstufigen Entscheidungsprozesses eine ganzheitliche Lösung zur Umstellung von konventionellen Fahrzeugflotten auf elektrische. In dem Entscheidungsprozess werden die technische, ökonomische und praktische Umsetzbarkeit von Elektromobilität untersucht. Angewendet auf eine konkrete Fahrzeugflotte mit Fahrstecken zwischen 13 und 77 km kann ein hohes Potential, herkömmliche Fahrzeuge zu ersetzen, festgestellt werden.

Potentiale und Herausforderungen. Die Möglichkeit, Elektromobilität in Fahrzeugflotten zu integrieren, wird von einem großen Teil der Literatur adressiert. Dabei spielen vor allem technische Aspekte wie die Reichweite eine Rolle. Methodisch wird überwiegend mit Umfragen und Feldtests gearbeitet sowie Datenerhebungen genutzt [KRZ16, BL16, Jo20, Gn13]. In [Ch17] und [KRZ16] werden verschiedene Sektoren auf deren Eignung zur Nutzung von Elektromobilität untersucht. Dazu werden unter anderem Fahrzeuganzahl, tägliche Strecken und Fahrleistung bewertet. Die Sektoren Bauwesen und Gesundheitswesen sind in dieser Hinsicht besonders gut für einen Wechsel zu Elektromobilität geeignet. Auf Basis von Flottendaten 16 verschiedener Unternehmen können [Be17] ebenfalls zeigen, dass ein Großteil der Strecken mit Elektrofahrzeugen zurückgelegt werden kann. [Iw21] testen die Energieeffizienz des elektrischen Lieferfahrzeugs Nissan eNV2000 unter realen Bedingungen. Die zurückgelegten Strecken sind in diesem Anwendungsfall relativ kurz. Dadurch kann auch hier die Reichweite von Elektrofahrzeugen nicht als Hindernis identifiziert werden. [QC21] hingegen zeigen, dass die Reichweitenlimitation eine Barriere für Logistikunternehmen darstellt, deren Standort außerhalb der Stadt liegt. Über die Analyse des Fahrverhaltens und eine Auswertung gefahrener Strecken wird in [Fi18] analysiert, ob Elektrofahrzeuge im Handwerksverkehr eingesetzt werden können. Es zeigt sich, dass die verminderte Reichweite der Fahrzeuge im Winter ein Problem darstellt.
Informationslücken über technische Eigenschaften von Elektrofahrzeugen können von [Di21] in einer empirische Studie unter Flottenmanagern festgestellt werden. Feldtests zeigen, dass über die Zeit die zurückgelegten Strecken zwischen zwei Ladungen gestiegen sind, was mit der gestiegenen Vertrautheit zu erklären ist [WHA14]. Auch [FW15] stellen fest, dass die Fahrer kein Bewusstsein über ihr eigenes Fahrverhalten haben, sodass das Elektrifizierungspotential vielfach unterschätzt wird. Im Unterschied zum Großteil der

Literatur betrachten [FW15] das Potential einer Elektrifizierung explizit für ländliche Unternehmen. Auswertungskriterien sind unter anderem der Zugang zu Ladeinfrastruktur, ökologisches und ökonomisches Potential, das Fahrverhalten und Informationen über die Fahrzeugflotte.
[Qu16] setzen zur Analyse des Potentials elektrische Lieferfahrzeuge in verschiedenen Städten ein. Im Ergebnis kann festgestellt werden, dass die Flexibilität beim Umstieg auf Elektromobilität abnimmt und ein höherer Planungsaufwand auf Grund der Ladezeiten entsteht. Die Abnahme der Flexibilität stellt gerade für kleine Unternehmen ein Problem dar. So benötigen kleine Unternehmen weiter herkömmliche Fahrzeuge, um Änderungen und Ausfälle kompensieren zu können [FW15, Fi18]. Auch [An21] zeigen, dass durch Ladezeiten sorgfältige Planungen erforderlich sind. Sie untersuchen die Herausforderungen der Elektrifizierung in der Last-Mile-Logistik. Die vorhersehbaren Routen stellen sich hier als Vorteil bei der Einführung heraus.

Motivation und Akzeptanz. Vorwiegend mit Umfragen und Feldtests wird untersucht, welche Faktoren die Akzeptanz beziehungsweise Motivation zur Nutzung von Elektromobilität beeinflussen. So ermitteln [GDW18] und [SSKP21] durch Umfragen unter Flottenmanagern beziehungsweise Kleinstunternehmern die Faktoren, die die Akzeptanz von Elektrofahrzeugen in kommerziellen Flotten beeinflussen. [Vu21] testen in einem Feldexperiment Elektrofahrzeuge im Unternehmensalltag. Auf diese Weise können die größten Hindernisse ermittelt und die Frage beantwortet werden, in wie weit verschiedene Barrieren die Einführung von Elektromobilität beeinflussen. [Mi19] vergeben Elektrofahrzeuge an Unternehmen mit dem Ziel, das Bewusstsein für Elektromobilität zu steigern.
Ein wichtiger Einflussfaktor ist das Wissen über und die Erfahrung mit Elektrofahrzeugen [SSKP21, Ka16]. Unternehmen, die bereits Erfahrungen mit Elektrofahrzeugen gesammelt haben, haben eine positivere Einstellung der Elektromobilität gegenüber [Ka16]. Auch die Begeisterung einzelner hat Einfluss auf die Frage, ob Elektrofahrzeuge im Unternehmen eingeführt werden sollten [GDW18]. Die Akzeptanz kann durch Feldexperimente, in denen Fahrzeuge von Nutzern getestet werden können, gesteigert werden [Vu21].
Identifizierte Gründe, Elektromobilität zu nutzen, sind Imagegewinn, Reduktion der laufenden Kosten [Ka16, Vu21] und ökologische Faktoren [GDS18, Vu21, Qu16, Ka16]. Als Hemmnis werden hohe Anschaffungskosten [Gi19, Bo18, Vu21] und die Reichweite [GDW18, Vu21] angesehen. Weitere Faktoren, mit denen Einfluss auf das Interesse an Elektromobilität genommen werden kann, sind finanzielle Anreize [Ja21, SSKP21].

Auswirkungen der Elektrifizierung. Der Einsatz von Elektromobilität hat unter anderem Auswirkungen auf die Kosten und die Emissionen des Unternehmens. Diese werden vor allem durch Fallstudien oder mathematisch-formale Modelle ermittelt [Bo18, Ma18, Ar20, Gh21]. Auswirkungen auf die Treibhausgasemissionen werden in [Ma18] als Entscheidungskriterium herangezogen, um Buslinien auf Elektromobilität umzustellen. In einem Anwendungsfall werden Buslinien untersucht, die auf Grund ökologischer Merkmale dazu geeignet sind, elektrisch zu fahren. [Ar20] vergleichen die Auswirkungen einer CO_2-Bepreisung auf Elektrofahrzeuge und konventionelle Fahrzeuge anhand realer Daten in der Last-Mile-Logistik.

Sie stellen fest, dass die eingeschränkte Reichweite den Kostenvergleich erschwert, da die Routen für Elektrofahrzeuge im Vergleich zu Routen, die mit konventionellen Fahrzeugen zurückgelegt werden, kürzer ausfallen.
[PP20] untersuchen Auswirkungen einer Elektrifizierung auf die Umwelt im städtischen Nahverkehr. Mittels einer Fallstudie können sie bestätigen, dass emissionsfreie Busse zur nachhaltigen Entwicklung und CO_2 Reduktion beitragen. Die Ergebnisse zeigen auch, dass der Wechsel zu Elektromobilität vor allem dann sinnvoll ist, wenn gleichzeitig auf erneuerbare Energien gesetzt wird.

Laden. Die Untersuchung des Ladeverhaltens der Fahrzeuge am Unternehmen ist ebenfalls ein wichtiges Themenfeld. Feldtests und mathematisch-formale Modelle sind hier methodisch vorherrschend [Te22, WBL19]. Nach [Mä21] ist das Laden aufgrund der zusätzlich einzuplanenden Zeit eine der größten Herausforderungen bei der Elektrifizierung. Sie stellen daher eine Funktion auf, um den Ladevorgang zu simulieren und die Ladedauer vorherzusagen. In Feldtests werden Strategien zum Umstieg auf Elektromobilität entwickelt. Es kann gezeigt werden, dass das Wissen über die Ladezeiten und Mengen eine hohe Bedeutung besitzt.
Bei der Untersuchung des Einsatzes von Elektromobilität im Transportbereich ist zu beobachten, dass meist ohne Ladestrategien geladen wird sondern immer dann, wenn die Möglichkeit zum Laden besteht [FFR20]. Als häufigster Ladezeitraum erweist sich 10-16 Uhr. In einem Zeitraum von 5 Jahren wurden hierzu 740 Elektrofahrzeuge in Unternehmen verschiedener Sektoren getestet.
Die Betrachtung des Ladevorgangs und der Ladeinfrastruktur zeigt, dass mehr Ladeinfrastruktur an einfach zugänglichen Orten benötigt wird, um Fahrten mit Ladestopps unterwegs zu ermöglichen [KHQ21]. [KHQ21] untersuchen die technischen, operativen und finanziellen Möglichkeiten verschiedener Ladestrategien in großen Fahrzeugflotten. Sie führen die Untersuchungen mit 39 Fahrzeugen über ein Jahr hinweg durch. Ergebnisse der Fallstudie zeigen, dass spontanes Laden ohne Reservierung der Ladesäule bevorzugt wird, die Verfügbarkeit jedoch vorher einsehbar sein sollte.

4 Fazit

In diesem Beitrag wurden 56 Paper aus den Jahren 2012-2022 analysiert, die sich mit Elektromobilität im betrieblichen Kontext beschäftigen. Über 50% der Paper sind ab 2019 erschienen. Es ist zu erkennen, dass sich die Literatur hauptsächlich mit dem Potential, das Elektromobilität bietet, sowie der Akzeptanz und den Auswirkungen beschäftigt. Die Unterstützung der Entscheidung und die Betrachtung des Ladens konnten ebenfalls als wichtige Themenfelder identifiziert werden.
Methodisch werden Feldtests und Fallstudien über alle Bereiche hinweg genutzt. Gerade bei der Ermittlung der Akzeptanz und des Potentials von Elektromobilität spielen Umfragen und Interviews eine wichtige Rolle. Zur Entscheidungsunterstützung und Untersuchung des Ladeverhaltens kommen vor allem mathematisch-formale Modelle zum Einsatz.

Im betrieblichen Kontext wird insbesondere der städtische Lieferverkehr untersucht, da er sich durch planbare Strecken und kurze Wege auszeichnet. Auffällig ist, dass hauptsächlich städtische Unternehmen betrachten werden, ländliche jedoch kaum, was bereits [Jo20] feststellt. Als relevantes Mittel zur Ermittlung des Potentials erweisen sich Fahrprofile. Weiterhin stellen ökologische Faktoren einen wichtigen Grund dar, auf Elektromobilität umzusteigen. Als wesentliche Erkenntnis aus den verschiedenen Forschungsfeldern erweist sich der Einfluss, den die Vertrautheit mit Elektromobilität auf die Nutzungsbereitschaft hat. Modelle sollten daher mit Piloteinführungen oder einfachen Möglichkeiten zum Testen von Elektromobilität verbunden werden.

Literaturverzeichnis

[An20] de Angelis, E.; Turrini, E.; Carnevale, C.; Volta, M.: Vehicle fleet electrification: impacts on energy demand, air quality and GHG emissions. An integrated assessment approach. IFAC-PapersOnLine, 53(2):16581–16586, 2020.

[An21] Anosike, Anthony; Loomes, Helena; Udokporo, Chinonso Kenneth; Garza-Reyes, Jose Arturo: Exploring the challenges of electric vehicle adoption in final mile parcel delivery. International Journal of Logistics Research and Applications, S. 1–25, 2021.

[Ar20] Arroyo, Jose L.; Felipe, Ángel; Ortuño, M. Teresa; Tirado, Gregorio: Effectiveness of carbon pricing policies for promoting urban freight electrification: analysis of last mile delivery in Madrid. Central European Journal of Operations Research, 28:1417–1440, 2020.

[Be15] Berndt, Rudiger; Schellenberg, Sebastian; Eckhoff, David; German, Reinhard: A Lightweight Simulation Framework for the Analysis of Electrified Vehicle Fleets. 2015 IEEE 81st Vehicular Technology Conference (VTC Spring), S. 1–5, 2015.

[Be17] Betz, Johannes; Hann, Moritz; Jager, Benedikt; Lienkamp, Markus: Evaluation of the Potential of Integrating Battery Electric Vehicles into Commercial Companies on the Basis of Fleet Test Data. 2017 IEEE 85th Vehicular Technology Conference (VTC Spring), 2017.

[BL16] Betz, Johannes; Lienkamp, Markus: Approach for the development of a method for the integration of battery electric vehicles in commercial companies, including intelligent management systems. Automotive and Engine Technology, 1:107–117, 2016.

[Bo18] Bousse, Yannick; Corazza, Maria Vittoria; Arriaga, Diego Salzillo; Sessing, Gerhard: Electrification of Public Transport in Europe: Vision and Practice from the ELIPTIC project. 2018 IEEE International Conference on Environment and Electrical Engineering, 2018.

[Br15] vom Brocke, Jan; Simons, Alexander; Riemer, Kai; Niehaves, Bjoern; Plattfault, Ralf; Cleven, Anne: Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research. Communications of the Association for Information Systems, 37(9):205–224, 2015.

[BVM86] Brans, J.P.; Vinke, Ph.; Mareschal, B.: How to select and how to rank projects: The Promethee method. European Journal of Operational Research, 24:228–238, 1986.

[Ch17] Christensen, Linda; Klauenberg, Jens; Kveiborg, Ole; Rudolph, Christian: Suitability of commercial transport for a shift to electric mobility with Denmark and Germany as use cases. Research in Transportation Economics, 64:48–60, 2017.

[CK12] Chlebišová, Eva; Kyzeková, Jana: The Information Sources for the Company Decision Making Process by Buying New Company Vehicles With Regard to Electric Vehicles. Acta academica karviniensia, 12(2):43–52, 2012.

[Di] Die Bundesregierung: Klimaschonender Verkehr https://www.bundesregierung.de/breg-de/themen/klimaschutz/klimaschonender-verkehr-1794672. Zuletzt Aufgerufen: 02.06.21.

[Di21] Di Foggia, Giacomo: Drivers and challenges of electric vehicles integration in corporate fleet: An empirical survey. Research in Transportation Business & Management, 41:3–25, 2021.

[FFR20] Fieltsch, Patrick; Flämig, Heike; Rosenberger, Kerstin: Analysis of charging behavior when using battery electric vehicles in commercial transport. Transportation Research Procedia, 46:181–188, 2020.

[Fi18] Figenbaum, Erik: Can battery electric light commercial vehicles work for craftsmen and service enterprises? Energy policy, 120:58–72, 2018.

[Fr] Fraunhofer IAO: Fraunhofer IAO elektromobilisiert Fuhrparkflotten https://www.muse.iao.fraunhofer.de/de/projekte/www-elektromobilisiert-de.html. Letzter Zugriff: 02.08.2021.

[FW15] Fornahl, Dirk; Wernern, Noreen: New Electric Mobility in Fleets in the Rural Area of Bremen/Oldenburg. In (Leal Filho, Walter; Kotter, Richard, Hrsg.): E-Mobility in Europe Trends and Good Practice, S. 237–251. Springer, 2015.

[Ga21] Galati, Antonino; Giacomarra, Marcella; Concialdi, Paola; Crescimanno, Maria: Exploring the feasibility of introducing electric freight vehicles in the short food supply chain: A multi-stakeholder approach. Case Studies on Transport Policy, 9(2):950–957, 2021.

[GDS18] Globisch, Joachim; Dütschke, Elisabeth; Schleich, Joachim: Acceptance of electric passenger cars in commercial fleets. Transportation Research Part A: Policy and Practice, 116:122–129, 2018.

[GDW18] Globisch, Joachim; Dütschke, Elisabeth; Wietschel, Martin: Adoption of electric vehicles in commercial fleets: Why do car pool managers campaign for BEV procurement? Transportation Research Part D: Transport and Environment, 64:122–133, 2018.

[Gh21] Ghandriz, Toheed; Jacobson, Bengt; Islam, Manjurul; Hellgren, Jonas; Laine, Leo: Transportation-Mission-Based Optimization of Heterogeneous Heavy-Vehicle Fleet Including Electrified Propulsion. Energies, 14:1–43, 2021.

[GHR21] Gorges, Tobias; Holz-Rau, Christian: Transition of mobility in companies - A semi-systematic literature review and bibliographic analysis on corporate mobility and its management. Transportation Research Interdisciplinary Perspectives, 11:1–10, 2021.

[Gi19] Giacomarra, Marcella; Tulone, Antonio; Crescimanno, Maria; Galati, Antonino: Electric mobility in the Sicilian short food supply chain. Studies in Agricultural Economics, 121(2):84–93, 2019.

[Gn13] Gnann, Till; Haag, Michael; Plotz, Patrick; Wietschel, Martin: Market potential for electric vehicles in the German commercial passenger transport sector. EVS27 International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium, S. 1–10, 2013.

[HE20] Hamurcu, Mustafa; Eren, Tamer: Electric Bus Selection with Multicriteria Decision Analysis for Green Transportation. Sustainability, 12:1–19, 2020.

[İÇK21] İmre, Şükrü; Çelebi, Dilay; Koca, Funda: Understanding barriers and enablers of electric vehicles in urban freight transport: Addressing stakeholder needs in Turkey. Sustainable Cities and Society, 68, 2021.

[IL19] Islam, Adrita; Lownes, Nicholas: When to go electric? A parallel bus fleet replacement study. Transportation Research Part D: Transport and Environment, 72:299–311, 2019.

[Iw19] Iwan, Stanislaw; Allesch, Juergen; Celebi, Dilay; Kijewska, Kinga; Hoé, Mette; Klauenberg, Jens; Zajicek, Juergen: Electric mobility in European urban freight and logistics – status and attempts of improvement. Transportation Research Procedia, 39:112–123, 2019.

[Iw21] Iwan, Stanisław; Nürnberg, Mariusz; Jedliński, Mariusz; Kijewska, Kinga: Efficiency of light electric vehicles in last mile deliveries – Szczecin case study. Sustainable Cities and Society, 74:1–10, 2021.

[Ja21] Jaller, Miguel; Pineda, Leticia; Ambrose, Hanjiro; Kendall, Alissa: Empirical analysis of the role of incentives in zero-emission last-mile deliveries in California. Journal of Cleaner Production, 317, 2021.

[Jo20] Jones, Andrew; Begley, Jason; Berkeley, Nigel; Jarvis, David; Bos, Elizabeth: Electric vehicles and rural business: Findings from the Warwickshire rural electric vehicle trial. Journal of Rural Studies, 79:395–408, 2020.

[Ju16] Juan, Angel; Mendez, Carlos; Faulin, Javier; de Armas, Jesica; Grasman, Scott: Electric Vehicles in Logistics and Transportation: A Survey on Emerging Environmental, Strategic, and Operational Challenges. Energies, 9(86):1–21, 2016.

[Ka16] Kaplan, Sigal; Gruber, Johannes; Reinthaler, Martin; Klauenberg, Jens: Intentions to introduce electric vehicles in the commercial sector: A model based on the theory of planned behaviour. Research in Transportation Economics, 55:12–19, 2016.

[KDC21] Kucukoglu, Ilker; Dewil, Reginald; Cattrysse, Dirk: The electric vehicle routing problem and its variations: A literature review. Computers & Industrial Engineering, 161, 2021.

[KH14] Koetse, Mark J.; Hoen, Anco: Preferences for alternative fuel vehicles of company car drivers. Resource and Energy Economics, 37:279–301, 2014.

[KHQ21] Kin, Bram; Hopman, Meike; Quak, Hans: Different Charging Strategies for Electric Vehicle Fleets in Urban Freight Transport. Sustainability, 13(23):1–18, 2021.

[KIM19] Kijewska, Kinga; Iwan, Stanislaw; Małecki, Krzysztof: Applying Multi-Criteria Analysis of Electrically Powered Vehicles Implementation in Urban Freight Transport. Procedia Computer Science, 159:1558–1567, 2019.

[KKR13] Koetter, Falko; Klausmann, Florian; Renner, Thomas: Potenzialermittlung für Integration von Elektrofahrzeugen in Fuhrparkflotten. Informatik-Spektrum, 36(1):35–45, 2013.

[KO16] Koetter, Falko; Ostermann, Julien: Real-Time Schedule Optimization in Shared Electric Vehicle Fleets. Proceedings of the International Conference on Vehicle Technology and Intelligent Transport Systems, 2016.

[Kr21] Krawiec, Krzysztof: Vehicle Cycle Hierarchization Model to Determine the Order of Battery Electric Bus Deployment in Public Transport. Transport Problems, 16(1):99–112, 2021.

[KRZ16] Klauenberg, Jens; Rudolph, Christian; Zajicek, Jürgen: Potential Users of Electric Mobility in Commercial Transport – Identification and Recommendations. Transportation Research Procedia, 16:202–216, 2016.

[LMV16] Lebeau, Philippe; Macharis, Cathy; Van Mierlo, Joeri: Exploring the choice of battery electric vehicles in city logistics: A conjoint-based choice analysis. Transportation Research Part E: Logistics and Transportation Review, 91:245–258, 2016.

[Ma16] Margaritis, Dimitris; Anagnostopoulou, Afroditi; Tromaras, Alkiviadis; Boile, Maria: Electric commercial vehicles: Practical perspectives and future research directions. Research in Transportation Business & Management, 18:4–10, 2016.

[Ma18] May, Nadine: Local environmental impact assessment as decision support for the introduction of electromobility in urban public transport systems. Transportation Research Part D: Transport and Environment, 64:192–203, 2018.

[Mä21] Märzinger, Thomas; Wöss, David; Steinmetz, Petra; Müller, Werner; Pröll, Tobias: Novel Modelling Approach for the Calculation of the Loading Performance of Charging Stations for E-Trucks to Represent Fleet Consumption. Energies, 14(12):1–15, 2021.

[Ma22] Malladi, Satya S.; Christensen, Jonas M.; Ramírez, David; Larsen, Allan; Pacino, Dario: Stochastic fleet mix optimization: Evaluating electromobility in urban logistics. Transportation Research Part E: Logistics and Transportation Review, 158, 2022.

[MBC14] Melo, Sandra; Baptista, Patricia; Costa, Álvaro: Comparing the Use of Small Sized Electric Vehicles with Diesel Vans on City Logistics. Procedia - Social and Behavioral Sciences, 111:350–359, 2014.

[Mi19] Mielzarek, Janine: Mobility Turnaround, Air Pollution Control and Secure the Budget - Electric Mobility in the Urban Area of Tension. Electric Vehicles International Conference, 2019.

[Na21] Napoli, Giuseppe; Polimeni, Antonio; Micari, Salvatore; Dispenza, Giorgio; Antonucci, Vincenzo; Andaloro, Laura: Freight distribution with electric vehicles: A case study in Sicily. Delivery van development. Transportation Engineering, 3:1–20, 2021.

[Os14] Ostermann, Julien; Renner, Thomas; Koetter, Falko; Hudert, Sebastian: Leveraging Electric Cross-Company Car Fleets through Cloud Service Chains: The Shared E-Fleet Architecture. 2014 SRII Global Conference, S. 290–297, 2014.

[Pa21] Patella, Sergio Maria; Grazieschi, Gianluca; Gatta, Valerio; Marcucci, Edoardo; Carrese, Stefano: The Adoption of Green Vehicles in Last Mile Logistics: A Systematic Review. Sustainability, 13(1):1–29, 2021.

[PP18] Papathanasiou, Jason; Ploskas, Nikolaos: Multiple Criteria Decision Aid Methods, Examples and Python Implementations, Jgg. 136. Springer International Publishing AG, 2018.

[PP20] Pietrzak, Krystian; Pietrzak, Oliwia: Environmental Effects of Electromobility in a Sustainable Urban Public Transport. Sustainability, 12(3):1–21, 2020.

[QC21] Qasim, Muhammad; Csiszar, Csaba: Major Barriers in Adoption of Electric Trucks in Logistics System. Promet - Traffic&Transportation, 33(6):833–846, 2021.

[Qu16] Quak, Hans; Nesterova, Nina; van Rooijen, Tariq; Dong, Yanjie: Zero Emission City Logistics: Current Practices in Freight Electromobility and Feasibility in the Near Future. Transportation Research Procedia, 14:1506–1515, 2016.

[Ro22] Rosenberger, Kerstin; Tapia, Rodrigo; Friedrich, Hanno; Flämig, Heike: Estimating the potential of electric mobility in commercial transport considering the availability of charging infrastructure – A behavioural model analysis for the city of Hamburg. Research in Transportation Business & Management, 2022.

[Si14] Sierzchula, William: Factors influencing fleet manager adoption of electric vehicles. Transportation Research Part D: Transport and Environment, 31:126–134, 2014.

[SSKP21] Skowrońska-Szmer, Anna; Kowalska-Pyzalska, Anna: Key Factors of Development of Electromobility AMONG Microentrepreneurs: A Case Study from Poland. Energies, 14:1–25, 2021.

[Te22] Tepe, Benedikt; Figgener, Jan; Englberger, Stefan; Sauer, Dirk Uwe; Jossen, Andreas; Hesse, Holger: Optimal pool composition of commercial electric vehicles in V2G fleet operation of various electricity markets. Applied Energy, 308:1–18, 2022.

[TKT16] Teoh, Tharsis; Kunze, Oliver; Teo, Chee-Chong: Methodology to Evaluate the Operational Suitability of Electromobility Systems for Urban Logistics Operations. Transportation Research Procedia, 12:288–300, 2016.

[Ts20] Tsakalidis, Anastasios; Krause, Jette; Julea, Andreea; Peduzzi, Emanuela; Pisoni, Enrico; Thiel, Christian: Electric light commercial vehicles: Are they the sleeping giant of electromobility? Transportation research. Part D: Transport and Environment, 86, 2020.

[Vu21] Vuichard, Pascal: Electrifying the company car: Identifying hard and soft barriers among fleet managers in Switzerland. Energy Research & Social Science, 77:1–11, 2021.

[Wą17] Wątróbski, Jarosław; Małecki, Krzysztof; Kijewska, Kinga; Iwan, Stanisław; Karczmarczyk, Artur; Thompson, Russell: Multi-Criteria Analysis of Electric Vans for City Logistics. Sustainability, 9:1–34, 2017.

[Wa18] Wang, M.; Thoben, K.-D.; Bernardo, M.; Daudi, Morice: Diversity in Employment of Electric Commercial Vehicles. Logistics Research, S. 1–21, 2018.

[WBL19] Waclaw, Adam; Betz, Johannes; Lienkamp, Markus: Techno-Economical Implementation of Holistic Electromobility Solutions in Commercial Companies. Fourteenth International Conference on Ecological Vehicles and Renewable Energies (EVER), 2019.

[WHA14] Wikström, Martina; Hansson, Lisa; Alvforsa, Per: Socio-technical experiences from electric vehicle utilisation in commercial fleets. Applied Science, 123:82–93, 2014.

Potentials of Bicycle Infrastructure Data Lakes to Support Cycling Quality Assessment

Johannes Schering [1], Jorge Marx Gómez[2] , Lena Büsselmann[3], Federico Alfaro[4], Jan Stüven[5]

Abstract: A data-driven quality assessment of bicycle infrastructure is necessary in times of crisis to support the decision-making process in cycling promotion. The INFRASense project was initiated to support the scoring of bike paths by providing new crowdsourcing data that is combined with other relevant data sources (traffic amount, accidents, citizen reportings etc.). The storage and processing of heterogeneous bike infrastructure data may be a challenge. With its flexibility a Data Lake could be an alternative to the traditional Data Warehouse. In the first step the paper gives an overview about data-driven initiatives in the use-case of bike infrastructure quality assessment and the recently started research project INFRASense. We will provide an overview about data sources that may potentially be included into the data driven quality assessment. Big Bicycle Data is available in many different structures and formats (CSV, XML, SHP etc.). In the second step the concepts of Data Lake and Data Warehouse are introduced. The benefits and weaknesses of these two solutions are shown followed by a discussion about which one of these is the best concept for storage, processing, and analysis of heterogeneous bicycle infrastructure data. In the last step we are providing an outlook how an efficient bicycle infrastructure data management system could be implemented.

Keywords: Bicycle Data, Data Lake, Data Warehouse, Bicycle Infrastructure, Quality Assessment

1 Introduction

In the past few years, a growing interest in cycling promotion can be perceived. The Corona pandemic has accelerated this trend. People more and more preferred individual means of transport such as bicycles or cars in order to maintain separation from other people. Especially the private car became even more popular with a high “feel good” factor [Ei21]. The city of Vienna that is traditionally famous for public transport

[1] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, johannes.schering@uni-oldenburg.de, https://uol.de/vlba/personen/mitarbeiterinnen/johannes-schering
[2] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, jorge.marx.gomez@uni-oldenburg.de
[3] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, lena.buesselmann@uni-oldenburg.de
[4] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, federico.alfaro@uni-oldenburg.de
[5] University of Oldenburg, Department of Business Informatics VLBA, Ammerländer Heerstr. 114-118, Oldenburg, 26129, jan.stueven@uni-oldenburg.de, https://uol.de/jorge-marx-gomez

doi:10.18420/inf2022_66

perceived a decrease of traffic participants in bus, tram etc. by 10 percent in the first year of the Corona pandemic [Sc22]. Due to multiple societal factors, it would be beneficial however, if more people were to change their mobility behaviour in favour of using the bike. For example, the use of active methods of transportation could provide health benefits, reduce operating costs of vehicles and reduce congestion [Li22]. The recent developments in energy and fuel costs have increased the magnitude of these benefits. To motivate more people to use the bike as a form of transport, the underlying biking infrastructure needs to be improved. It is proved by scientific research that better cycling infrastructure can encourage more cycling [HuOH14].

In the case of German cycling infrastructure, it is often the responsibility of municipalities to invest in and maintain cycling infrastructure. However, according to the German Association of Towns and Municipalities, German cities will experience a deficit of 19.4 billion € until 2024 as a consequence of the Corona crisis. At the same time, municipalities can expect increasing costs due to the consequences of the Ukraine war which can include care for refugees or energy costs [DSt22]. Therefore, municipalities need to focus on cost efficient expenditures with regards to cycling infrastructure. In order to facilitate this, municipalities require information regarding the existing cycling infrastructure. At this moment this information is not available or of poor quality. The available information is mainly limited to surveys, traffic countings, accident analysis and traffic observations [MT10]. In addition, the perspective of the cyclists is not considered in the planning process. Experts in bicycle infrastructure planning need standardized assessment criteria which are based on reliable data that reflects real problems in bicycle use such as waiting times and the perspective of the cyclists. A quality assessment based on crowdsourcing data could support the planning process of the bike infrastructure and the efficient investment in the network. As we will see in the following, some data sources that may support the bike infrastructure quality assessment are already available. From the data processing perspective, the question is how best to store and process heterogeneous data sets from cycling infrastructure domain. It will be discussed in this contribution whether a Data Lake may be an alternative to the traditional Data Warehouse.

2 Data Driven Quality Assessment

2.1 State of the Art

The existing approaches to score bike infrastructures are described in the following. The overview helps to understand the research gap and the relevance of a recently started research project that is further described below. Many municipalities are providing platforms for citizen communication. Inhabitants and visitors have the opportunity to get in touch with the local administration. As an example, the city of Osnabrück provides the

EreignisMeldeSystem Osnabrück (EMSOs, *Event Reporting System Osnabrueck*)[6] portal. Users can report incidents and choose between different categories (roads, garbage disposal, green areas etc.), mark the location on the map and write a text to describe the event or shortcoming. In terms of cycling, the *Allgemeiner Deutscher Fahrrad-Club* (ADFC, *German Bicycle Club*) provides platforms to enable cyclists to send information about critical bike infrastructure and (perceived) shortcomings. At the *ADFC Mängelmelder* ((*ADFC Deficiency indicator*)[7] for the city of Bremen cyclists may choose a location, select different (sub-)categories (traffic lights, construction sites, road damages, illegal parking etc.) and write a description. The limitation here is that the reports only reflect the subjective perceptions of individual persons. The feedback is not suitable for deriving general and comparable information about cycling quality.

The *Happy Bike Index* provides a first data driven and objective quality assessment for inner city intersections based on data from the municipal administrations.[8] In the current version the assessment considers (perceived) cycling safety regarding automobile traffic. The used indicators are the allowed maximum speed level and the daily traffic volume. At the moment, the beta version of the index does not consider construction conditions of the bike infrastructure like width, surface quality, or state. These data sets should be considered as soon these are available. Still missing are real problems on the bike trips that are detected by cycling people (e.g. waiting times at intersections).

CROW Fietsberaad which is a cycling expert committee from the Netherlands, has made a suggestion of how bike paths can be classified into quality levels on a range from A (=very good) to E (=very bad). The central aspect is the bike path width. As a result of different studies this aspect was identified as key for comfort and traffic safety. It becomes more and more relevant with the uptake of pedelec sells – because there are more and more bicycles with different speed levels on the road. If there is enough space, there is room for overtaking or evasion movements. As another aspect, the traffic amount is considered. If motorized bikes are allowed on the bike paths this has also an influence on the quality assessment. The amount of cargo and racing bikes should also be considered [VB21]. Similar to the Happy Bike Index, data of real bike trips and cyclist feedback is not considered in the methodology.

The creation of a holistic quality assessment of the bicycle infrastructure may be a challenge as we have different types of bike paths and alignment of these. The cycling network can be differentiated into advisory bike lanes and bike lanes on the road, one- and two-direction bike lanes, mixed pedestrian and bike path, bicycle boulevards that prioritize cyclists and road segments where cyclists are riding on the road together with motorized traffic without any protection [Ko13]. The *Forschungsgesellschaft für Straßen- und Verkehrswesen* (FGSV, *Research Association for Road and Traffic Systems* in Germany) that brings together various experts from the bicycle traffic planning

[6] https://geo.osnabrueck.de/emsos/
[7] https://adfc-bremen.mängelmelder.de/
[8] https://fixmyberlin.de/zustand

domain in Germany is responsible to define new guidelines and standards for planning. In 2021, the expert team published a new guideline regarding the consistent and standardized assessment of bike paths to create a solution for the challenge described above. The *Hinweise zur einheitlichen Bewertung von Radverkehrsanlagen* (H EBRA, *Suggestions for the Consistent Assessment of Bike Paths*) gives an overview about the most important criteria and the weighting of these from the planning perspective. There are three dimensions as assessment aspects: 1) alignment of the bike path, 2) state of the bike path and 3) attractiveness. The alignment is related to a) motorized traffic and b) pedestrian traffic. The roads are classified to congestion classes (traffic amount). The pedestrian traffic is considered when roads have a lot of shops, a lot of pedestrians in need of protection (children, handicapped people etc.), important connections for bicycle traffic, steep declivity, house and property entrances at narrow side paths, bus and tram stops, high number of pedestrians in relation to side path width etc. The state of the bike path is divided into shortcomings at a) specific locations (potholes, bollards, root breakup etc.), b) over the whole length of the path that causes time losses (surface types with high vibration level, missing markings etc.), c) junctions (missing bicycle traffic light signaling, long waiting times at red lights etc.) and d) overtakings (bike path width, illegal parking cars etc.). The attractiveness is divided into travel experience for cyclists (fresh air, green environment, low sound emissions etc.) and the perceptibility of the travel experience for cyclists (only fulfilled when whole quality level is good, pedestrian and/or bicycle traffic amount is not to high etc.) [FGSV21].

The H EBRA guideline includes an overview about demanded data sources. Geographic Information Systems (GIS) could give insights to length, inclines and alignment of bike paths. Construction criteria such as junction type, bike path width, surface state and type need to be taken into account. The traffic amount can be measured by day. Maximum speed levels for motorized vehicles may be considered. Bicycle counters deliver daily cyclist numbers at local spots of the city. When the side path is used by cyclists and pedestrians the amount of people that are walking and related conflict situations are important. The different types of shortcomings and the travel experience for cyclists that are described above should also be part of the data sources. Especially when it comes to problems on real bike trips it remains unclear what kind of data may be used [FGSV21]. This kind of “real cyclists” data is also not considered in the Dutch approach of CROW Fietsberaad described above.

2.2 INFRASense

To find new solutions to the challenges described above, a new bicycle data driven research project INFRASense[9] was initiated. The main goal is the development of a sensor-based quality assessment of the bike infrastructure which is based on crowdsourcing data and consists of a software tool including a mobile app for the analysis and the definition of quality levels (A = very good to F = very bad). The

[9] www.infrasense.de

software tool *Bicycle Lane Quality Evaluation* (BIQE) will be a new basis for the decision making in the bike infrastructure planning process. The involvement of real cyclists is a very important aspect to create an assessment tool that reflects the user perspective. At least 1.000 citizens will participate in the data collection using a specialized sensor box that measures GPS, acceleration (which allows the analysis of vibrations on the bicycle) and environmental data directly on the bike. Further information about the bike sensors can be found in Schering et al. [Sc21]. Another 750 citizens will use the user interface of the software tool to provide feedback about the quality assessment to the project. With the user interface citizens can send information about individual bike rides and quality assessment The feedback will be used to make the quality assessment more accurate. Machine Learning methods for the automatized classification of the bike paths will be applied. As a result of the project we will have a system that consists of a software solution, bike sensors and analysis methods that provide crowdsourcing data-based information about cycling quality. The provided Key Performance Indicators (waiting and loss times, surface qualities, traffic flow, weather influence, comparison of bad bike paths quality with the number of accidents etc.) and visualizations will support the automatic quality assessment.

The development requires a comprehensive data management system in the background. To process all the external and internal data sets an architecture for data integration, -analysis and -provision is needed. The following working steps include the construction of a database, the implementation of interfaces to municipal GIS or open data portals, automated sensor export, etc. The data management should enable the data analysis of several criteria relevant to cycling quality.

3 Data Sources and Approaches for Data Processing

3.1 Data sources

Many different data sources are needed to enable the quality assessment of bicycle infrastructure. On the one hand, the bike sensor is very important for obtaining user-specific information about the quality of a bike trip. The data can be used, for example, to answer questions about where cyclists have to stop particulary often or at which intersection waiting times are very long. On the other hand, external data sources contain data on the bike path network (material, path type, width), traffic planning (bike counting, traffic flow, occupancy rate etc.), shape files of municipal GIS (bus stops, neighbourhood borders etc.) and accident data from the police. In the following, Tab. 1 provides an overview of the available data sources that may be used in the quality assessment.

Provider	Contents	Data Format	Interface
CoSynth GmbH (internal)	Bike sensor data: Accelerometer, GPS etc.	CSV	API (regular)
Project internal data collection	Camera data, images	MP4	Reading out (per experiment)
Planungsbüro VIA (internal)	Bike path width	Excel	Will be collected once and transmitted by email
City administration (Department of traffic guidance)	Locations of intersections	PDF	e-mail (once)
City administration (Department of traffic guidance)	Traffic volume, occupancy rate	CSV	API (daily)
City administration (Department of Mobility)	Temporary and permanent Bike countings	PDF, CSV	Webportal (Download, e.g. per month), email
City administration (Department of Mobility)	List of bike paths (obligatory use), bridges, snow-plowing service	PDF	e-mail (once)
City administration (Department of Maintenance)	Road register	Excel	e-mail (once)
City administration (Department of communication)	Reportings about shortcomings	CSV	*not defined yet
City administration (Department of Geoinformation and Statistics)	Geoinfornation: Cycling network, bike parking facilities, bus stops, topography, districts, road classification …	SHP/SHX/ SBN/CPG/ DBF/ PRJ	e-mail (once)
Police Inspection	Historical bicycle accident data	Excel	e-mail (once per year)
Visual Crossing	Weather data	CSV	API (daily)

Tab. 1: External data sources

3.2 Data Warehouses

As we learnt from the overview above, the project has to deal with heterogeneous data sources. These consist of CSV, XML, Word, PDF, SHP or GeoJSON files. The different data formats have the effect that data management must be able to handle structured, semi-structured and unstructured data sets. The heterogeneity could be a challenge in the data integration process for traditional data management and analysis systems. There are different concepts and technologies for data storage. In the past, files, data bases, data warehouses, cloud systems and data lakes have been proposed [ZKBG21].
The Data Warehouse can be defined as an integrated and historical storage system [SD21]. It can be used for data analysis purposes and was developed to support the decision-making processes in businesses [Kr13]. Relevant data sets are extracted from different source systems, are transformed into the schema of the target Data Warehouse and are saved in it (in batches, in real time etc.) [DS16]. The Data Warehouse has four central characteristics. According to a definition of Bill Inmon (1996) who is known as the father of Data Warehousing it is subject oriented (e.g., biking), integrated (transforming data from different sources into a consistent format), non-volatile (data in the warehouse should not change anymore) and time-variant (changes over time may be observed) [In96]. All data sets with relation to the same topic will be interconnected and standardized as part of the Extract, Transform, Load (ETL) process. The data sources can be provided by different operational systems and may have variations in file layouts, code representation and header names [Po10]. As part of the ETL process the data sets are transformed and transferred at defined time intervals. The historical data will be saved and will not be changed anymore. The storage of historical data is a key characteristic of the Data Warehouse [Po10].

3.3 Data Lake

A potential alternative for the storage and integration of big and heterogeneous data could be a Data Lake, which functions as a raw data repository to store all data sets in any format [Ch20]. The Data Lake concept was introduced in 2010 by James Dixon [Ma17, MT16]. Since then, the concept has become well established in the industry. In the research domain, however, acceptance of the Data Lake is still in the very beginning [Zo21]. The implementation of Data Lake technologies for processing and storage of cycling data was not considered in research. Depending on the data sets of the Data Lake and the use case scenario there are different requirements regarding a Data Lake architecture [Gi21]. As bicycle data is still a quite young research domain it is not clear how suitable Data Lake approaches from the literature could be applied in the context of the bike infrastructure analysis and quality assessment.

A Data Lake can be defined as a central location for data storage. Data can be stored and applied in the Data Lake in any format [SD21, TR21]. In addition, data sets can be added to the Data Lake without changing its data structure [MT16] to enable an analysis

without any pre-defined use cases. The Data Lake is a collection of different data sets [TN20]. Data import may be realized by any method [Ma17]. An important difference compared to the Data Warehouse is, that the ETL process will not be applied directly. In contrast the data sets will be extracted from their origin source and will be stored in the Data Lake. In the next step data can be transformed in its structure. This process may be called Extract, Load, Transform (ELT) [Ka18]. Information can be extracted from the Data Lake as part of data processing and analysis [Ma17]. The Data Lake may deal with structured, semi-structured and unstructured data. It is often applied in the case of a fast-growing amount of big unstructured data sets, which makes it different to the Data Warehouse [TN20].

The application of the Data Lake brings new challenges for storage, processing and search of data [TR21]. An important challenge relates to the metadata of the data sources. There may not be a complete data catalog for the data sources in the Data Lake. In addition, the data possibly does not contain metadata relevant to its use or utilizes different formats to describe its metadata [Na18]. Finding data may be difficult because no global schema exists [TR21]. Without the appropriate structure and organization of the data, a Data Lake can turn into a so-called Data Swamp, which limits the use of the data [KW18, TR21, TN20]. Further problems arise in the implementation of Data Lakes as there is no universal data lake architecture that covers all necessary aspects [Gi19].

3.4 Differentiation between Data Lake and Data Warehouse

The concepts of the Data Lake and the Data Warehouse are often compared to each other as a system to store data for data analysis [TR21]. However, the data processing approaches are very different. As described above, the Data Warehouse relies on the ETL, the Data Lake on the ELT process. The Data Lake imports the raw data as it is and the structure is only transformed when there is demand for that. That makes the concept more flexible and agile [KW18]. The Data Warehouse is characterized by less flexibility. The schema of the Data Warehouse is developed and determined before the data is loaded. The Data Warehouse is efficient for the storage and the processing of structured data from operative systems [SD21]. Semi- and unstructured data may be a challenge because of its heterogeneity and complexity [Gi19, ML16].

The Data Lake may contribute to find a solution to these challenges which arise from the growing amount of available (big) data. There is a broad consensus in the literature that the architectures currently available to enable the integration of this data is primarily the Data Lake [CCWQ20]. As we already learned from the data overview in Tab. 1 above, in the specific use case of bike infrastructure there a quite a lot of data sets available. Bike infrastructure already falls to the category of Big (Bicycle) Data simply because of the volume involved [Ro16]. Therefore, it can be assumed that the Data Lake approach may be a suitable solution for the realization of a holistic system for the quality assessment of the bike infrastructure.

The approach that is discussed in this contribution allows further data processing regarding the quality of the bicycle infrastructures. A lot of measures that may contribute to a better bike path network are imaginable. Based on the detected waiting times at intersections at frequently used roads the traffic signalling may be adjusted. Surface types can be changed (e.g. to asphalt) or the state of the surface can be renewed at locations with a high number of bumps on the road. Evasion movement hotspots can be an indicator that space is not sufficient (e.g. because of a mixed space scenario with pedestrians). As a consequence, the bike path or side walk may be broadened. All these measures may support cyclists to avoid to have time losses and to make cycling more attractive. Traffic safety is another important factor whether people are using the bike or not. If the data processing reveals roads with many truck-bicycle accidents the bike path alignment could be changed (e.g. to a protected bike lane). In roads with many car accidents the allowed maximum vehicle speed level may be reduced from 50 to 30 km/h. If we identify locations where many people are complaining about parking violators, the number of parking spaces could be reduced. That will contribute to the availability of space, less conflicts with pedestrians and hopefully to a more liveable city.

4 Conclusion and Outlook

A data-driven quality assessment of bicycle infrastructure is necessary in times of crisis to support the decision-making process in cycling promotion. Existing approaches are limited to the subjective perceptions of cyclists or only take into account traffic safety or construction criteria data. The H EBRA guideline gives an orientation of how a holistic quality assessment could be realized. The INFRASense project was initiated to support this process by providing new crowdsourcing data that is combined with other relevant data sources. Since we have to deal with a huge amount of data to create a holistic quality assessment of road segments, the question arises what could be the most suitable solution for data storage and processing. Data Warehouses have a long tradition in this field as they are able to restructure and process huge amounts of data. For the development of the data driven quality assessment, a Data Lake could be an alternative. More flexibility is facilitated because it is able to handle heterogeneous.

The next step is to decide which ETL or ELT approach is best suited for processing of bicycle infrastructure data. The external data sources need to be combined with the bike sensor data to enable more specific data analysis. Potential connections between the data sources could be established using the time stamps, geo coordinates and road names. For example, road lengths could be linked to sensor or accident data in order to calculate how many brakings, stops or accidents occur per meter on a specific road segment. These associated Key Performance Indicators support the bicycle infrastructure assessment process and will be discussed in further publications. Another requirement is that GIS data may be stored and processed as part of the data management.

One potential option for the implementation could be a Data Lake in combination with a

Data Warehouse. As described above, the Data Lake stores all the raw data sets regardless of their format and structure without changing them. The Data Lake has an in-build API functionality to enable the implementation of the ETL or ELT process. The API allows communication between raw data storage and ETL or ELT methodology. In addition to that a Data Warehouse could be implemented. In the Data Warehouse the data may be transformed into a specific format, taking into consideration the specific requirements of the domain experts in the field of traffic planning. More specific research questions can be answered as part of the data analysis. In any case, these results will be presented in further publications.

Acknowledgements

INFRASense is funded by the Bundesministerium für Digitales und Verkehr (BMDV, *German Federal Ministry of Digital and Transport*) as part of the mFUND program (project number 19F2186E) with a funding amount of around 1.2 Mio. Euro. As part of mFUND the BMDV supports research development projects in the field of data based and digital mobility innovations. Part of the project funding is the promotion of networking between the stakeholders in politics, business, administration and research as well as the publication of open data on the mCLOUD portal (which will be superseded by Mobilithek in 2022/23).

References

[BSS21] Bauer, M.; Sanchez, L.; Song, J.: IoT-Enabled Smart Cities: Evolution and Outlook. Sensors 21, Nr. 13, pp. 1-29, 2021.

[Ch20] Chiusano, S.; Cerquitelli, T.; Wrembel, R.; Quercia, D.: Breakthroughs on Cross-Cutting Data Management, Data Analytics, and Applied Data Science. Information systems frontiers 23, Nr. 1, pp. 1-7, 2020.

[DS16] Dhanda, P.; Sharma, N.: Extract Transform Load Data with ETL Tools. International Journal of Advanced Research in Computer Science 7, Nr. 3, pp. 152-160, 2016.

[DSt22] Deutscher Städte- und Gemeindebund: Kommunale Finanzen: Einführung, https://www.dstgb.de/themen/finanzen/, accessed 10/05/2021.

[Ei21] Eisenmann, C.; Kolarova, V.; Nobis, C.; Winkler, C.; Lenz, B.: DLR-Befragung: Wie verändert Corona unsere Mobilität? Verkehrsmittelnutzung, Einkaufs-, Arbeits- und Reiseverhalten, https://verkehrsforschung.dlr.de/de/news/dlr-befragung-wieveraendert-corona-unsere-mobilitaet, accessed 04/03/2021.

[FGSV21] Forschungsgesellschaft für Straßen- und Verkehrswesen (FGSV): Hinweise zur einheitlichen Bewertung von Radverkehrsanlagen (Ausgabe 2021). FGSV Verlag, Köln, 2021.

[Gi21] Giebler, C.; Gröger, C.; Hoos, E.; Eichler, R.; Schwarz, H.; Mitschang, B.: The Data Lake Architecture Framework: A Foundation for Building a Comprehensive Data Lake Architecture. In: Proceedings der 19. Fachtagung fuer Datenbanksysteme fuer

Business, Technologie und Web (BTW 2021), pp. 351-370, 2021.

[Gi19] Giebler, C.; Gröger, C.; Hoos, E.; Eichler, R.; Schwarz, H.; Mitschang, B.: Leveraging the Data Lake: Current State and Challenges. In: Ordonez, C.; Song, I.-Y.; Anderst-Kotsis, G.; Tjoa, A M.; Khalil, I. (ed.): Big Data Analytics and Knowledge Discovery. Springer Nature Switzerland AG, pp. 179–188, 2019.

[HuOH14] Hull, A.; O'Holleran, C.: Bicycle infrastructure: can good design encourage cycling? Urban, Planning and Transport Research. Routledge, 2(1), pp. 369–406, 2014. doi: 10.1080/21650020.2014.955210.

[In96] Inmon, W.H.: Building the Data Warehouse. John Wiley & Sons, Inc., New York, 1996.

[KW18] Khine, P.; Wang, Z.: Data lake: a new ideology in big data era. ITM Web of Conferences 17, pp. 3025, 2018.

[Ko13] Köhler, U.: Einführung in die Verkehrsplanung. Grundlagen, Modellbildung, Verkehrsprognose, Verkehrsnetze. Fraunhofer IRB Verlag, Stuttgart, 2013.

[Kr13] Krishnan, K.: Data Warehousing in the Age of Big Data. Morgan Kaufmann, Boston, 2013.

[Li22] Litman, T.: Evaluating Active Transport Benefits and Costs. Guide to Valuing Walking and Cycling Improvements and Encouragement Programs. Victoria Transport Policy Institute, Victoria, 2022.

[ML16] Madera, C.; Laurent, A.: The Next Information Architecture Evolution: The Data Lake Wave. In: 8th International Conference on Management of Digital EcoSystems, MEDES 2016, pp. 174–180, 2016.

[Ma17] Mathis, C.: Data Lakes. Datenbank-Spektrum 17, Nr. 3, pp. 289-293, 2017.

[MT10] Mettenberger, T.; Thiemann-Linden, J.: Bewertung der Radverkehrsqualität. Methoden der Wirkungskontrolle und stadtweite Fahrradbilanzen. Forschung Radverkehr, O-2, Deutsches Institut für Urbanistik –Difu-, Berlin, 2010.

[MT16] Miloslavskaya, N.; Tolstoy, A.: Big Data, Fast Data and Data Lake Concept. In: Procedia Computer Science 88, pp. 300-305, 2016.

[Na18] Nargesian, F.; Zhu, E.; Miller, R.J.; Pu, K.Q.; Arocena, P.C.: Data Lake Management: Challenges and Opportunities. In: Proceedings of the VLDB Endowment, Nr. 12, pp. 1986-1989, 2018.

[Po10] Ponniah, P.: Data Warehousing Fundamentals for IT Professional. 2nd ed. John Wiley & Sons Inc., Hoboken, New Jersey, 2010.

[Ro16] Romanillos, G.; Zaltz A.M.; Ettema, D.; De Kruijf, J.: Big Data and Cycling. Transport Reviews 36, Nr. 1, pp. 114–133, 2016.

[SD21] Sawadogo, P.; Darmont, J.: On Data Lake Architectures and Metadata Management. Journal of Intelligent Information Systems 56, Nr. 1, pp. 97-120, 2021.

[Sc21] Schering, J.; Janßen, C.; Kessler, R.; Dmitriyev, V.; Marx Gómez, J.; Stehno, C.; Pelzner, K.; Bankowsky, R.; Hentschel, R.: ECOSense and its preliminary findings: Collection and analysis of bicycle sensor data. In: Kamilaris, A.; Wohlgemuth, V.;

Karatzas, K.; Athanasiadis, I. (ed.). EnviroInfo 2020 Environmental Informatics – New perspectives in Environmental Information Systems: Transport, Sensors, Recycling. Adjunct Proceedings of the 34th EnviroInfo conference. Shaker Verlag, Düren, pp. 145-153, 2021.

[Sc22] Scherreiks, W.: Weiter zur smarten Stadt – Zukunftsplan Wien. VELOPLAN 01/22, pp. 18-27, 2022.

[TR21] Taniar, D.; Rahayu, W.: Data Lake Architecture. In Barolli, L.; Natwichai, J.; Enokido, T. (ed): Advances in Internet, Data and Web Technologies. Springer Nature Switzerland, pp. 344-357, 2021.

[TN20] Thomas, K..; Nair, P.: Data Lake: A Centralized Repository. International Research Journal of Engineering and Technology (IRJET) 07, Nr. 5, pp. 2978-2981.

[VB21] Veroude, B.; van Boggelen, O.: Discussienotitie - Actualisatie aanbevelingen voor de breedte van fietspaden. CROW Fietsberaad, Utrecht, 2021.

[Zo21] Zouari, F.; Boukadi, K.; Kabachi, N.; Ghedira-Guegan, C.: Data Management in the Data Lake: A Systematic Mapping. ACM International Conference Proceeding Series, pp. 280-284, 2021.

Mobilitätswende @ Lebensqualität (MobiLe) – Die Entwicklung eines qualitativen Modells zur Vermittlung wichtiger Wirkungszusammenhänge im komplexen System Verkehr

Herbert Brüning[1] Jorge Marx Gómez[2] Alexander Papke[3] Sreedhar Kokkarachedu[4]

Abstract: Das Forschungsvorhaben MobiLe zielt darauf ab, eine neuartige Entscheidungshilfe für verkehrsbezogene Entscheidungen zur Verfügung zu stellen, die das Verkehrssystem, zusammen mit weiteren Aspekten einer qualitativen, auf Nachhaltigkeit zielenden Stadtentwicklung in ihrem systemischen Zusammenspiel betrachtet. Dabei wird ein webbasiertes Planungstool entwickelt, das mit einer leicht nutzbaren Software die Kommunalpolitiker*innen in die Lage versetzt, die Komplexität und Interdependenzen des kommunalen Verkehrssystems besser zu verstehen und zu berücksichtigen. Dies ermöglicht es den Kommunalpolitiker*innen, die komplexen Auswirkungen verkehrswirksamer Entscheidungen eigenständig und frühzeitig abzuschätzen.

Keywords: Stadtentwicklung; webbasiertes Planungstool; kommunale Verkehrssysteme; kybernetisches Modell; Mobilitätswende

1 Einleitung:

Eine Mobilitätswende ist im Kontext einer nachhaltigen Entwicklung erforderlich [De18], aber bislang nur unzureichend auf den Weg gebracht [De21]. Mit einer Anderen, auf umweltfreundliche Mobilität ausgerichteten Gestaltung unseres Verkehrssystems [Co20] lassen sich viele positive Effekte erreichen [NH16] – vom Rückgang gesundheits- und umweltschädlicher Luft- und Lärmbelastungen über eine geringere Inanspruchnahme von Flächen und anderen Ressourcen bis hin zu den positiven gesundheitlichen Effekten einer aktiven Bewegung. Das passt in das Konzept der „Städte für Menschen“ [Ge15], die eine hohe Lebensqualität und damit Attraktivität aufweisen. Kopenhagen ist das Beispiel, an dem dieses Prinzip erstmalig angewendet wurde. Vieles spricht also für eine auf Nachhaltigkeit zielende Neuausrichtung der Mobilität. Das betrifft auf kommunaler Ebene, deren besondere Bedeutung für die Verkehrswende in der aktuellen Version der Deutschen Nachhaltigkeitsstrategie betont wird [Di21], insbesondere die Nahmobilität.

[1] Stadt Norderstedt, 22846 Norderstedt, Deutschland Herbert.bruening@norderstedt.de
[2] Carl von Ossietzky Universität Oldenburg, Wirtschaftsinformatik, 26129 Oldenburg, Deutschland jorge.marx.gomez@uol.de
[3] Stadt Norderstedt, 22846 Norderstedt, Deutschland Alexander.papke@nordersted.de
[4] Carl von Ossietzky Universität Oldenburg, Wirtschaftsinformatik, 26129 Oldenburg, Deutschland sreedhar.kokkarachedu@uni-oldenburg.de

doi:10.18420/inf2022_67

Eine besondere Herausforderung für die Mobilitätswende stellt die Komplexität des Verkehrssystems dar. Viele Menschen unterschätzen diese Komplexität, da für nahezu alle Menschen die Teilnahme am Verkehrsgeschehen alltäglich ist und darüber eine Vielfalt an eigenen Erfahrungen vorliegt. Das verleitet dazu, sich beim Thema Verkehr als Expertin bzw. Experten zu betrachten. Darin unterscheiden sich – in Norderstedt wie in vielen anderen Städten auch – die Entscheider*innen nicht grundsätzlich von der breiten Bevölkerung. Daher ist es für eine Mobilitätswende notwendig, das zumeist laienhafte Verständnis der Systemzusammenhänge durch „Laien bzw. Laiinnen“ in Entscheidungspositionen zu erhöhen. Dieses Verständnis muss so sicher sein, dass Systemzusammenhänge in die Bevölkerung hinein gut vermittelbar sind – sowohl durch die Politiker*innen als auch durch die Verwaltung. Die für eine Mobilitätswende nötigen Transformationsprozesse müssen einerseits von fachfremden Personen (die z.B. in der Politik Verantwortung tragen) und andererseits für fachfremde Personen (der Großteil der Bevölkerung) überzeugend argumentiert werden können. Zu diesem Zweck soll mit MobiLe eine passende Unterstützung entwickelt werden, die Zusammenhänge verdeutlicht und ebenso für bekannte wie auch für neue Elemente des kommunalen Verkehrssystems anwendbar ist. MobiLe stellt dabei einen wissensbasierten Ansatz als weitere Entscheidungsgrundlage bereit, die eine Mobilitätswende argumentativ unterstützt.

Wenn sich die Erwartungen in diesem Ansatz einer „Politikberatung“ erfüllen, dann nutzen die Politiker*innen aller Fraktionen MobiLe zur Vorbereitung ihrer politischen Beratungen und Beschlüsse. Sie geben alle von ihnen gewünschten Maßnahmen über die sieben Stellgrößen ein, schauen sich das Ergebnis der Simulation an und überprüfen es anhand ihrer politischen Ziele. Werden diese erfüllt, dient MobiLe als Absicherung der eigenen Position. Werden diese noch nicht erreicht, kann über die Nutzung der Analysefunktionen nachvollzogen werden, an welchen Stellen sich das System unerwartet verhält und wirkungsvoll beeinflussen lässt. Selbst wenn dieser Analyseschritt nicht (in jedem Einzelfall) genutzt wird, lassen sich durch Veränderungen an den Stellgrößen schnell neue Simulationen durchführen, sodass eine Optimierung auch spielerisch erreicht werden kann. Eine der Voraussetzungen für den Erfolg wird sein, dass die Anwendung dieses Werkzeugs schnell geht; die politische Arbeit ist für die ehrenamtlich tätige Kommunalpolitik auch so schon sehr zeitintensiv.

2 Ausgangslage

Die Entwicklung des kybernetischen, qualitativ ausgerichteten Modells findet im Vorhaben „Mobilitätswende @ Lebensqualität“ statt, das im Rahmen der MobilitätsWerkStadt 2025 vom BMBF aktuell bereits in der zweiten Phase gefördert wird. Grundlage der Entwicklung des Modells sind daher neben allgemeinen wissenschaftlichen und technischen Quellen auch die Ergebnisse aus der ersten Phase des Vorhabens.

2.1 Wissenschaftlicher und technischer Stand

Für eine qualitative Abbildung und Analyse verkehrlicher Zusammenhänge und Entwicklungen kann auf erste Vorarbeiten zurückgegriffen werden. So werden systemdynamische Sensitivitätsmodelle [Ve88][Ko11] bereits in Forschungsvorhaben der Verkehrs- und Raumplanung verwendet [Kö92][SRS02][Kl08][GH20]. Als Vorteile dieser Modelle werden der vergleichsweise geringe Datenaufwand und die hohe Nachvollziehbarkeit durch die Entscheidungsträger*innen genannt [Kl08][HWB19].

Bei einem Forschungsvorhaben wurde beispielsweise mit dem Sensitivitätsmodell ein Wirkungsgefüge „Nachhaltige Mobilität“ erarbeitet [Be05]. Dieses ist in einer Dissertation zum Systemmodell „Stadtverkehr“ weiterentwickelt und veröffentlicht worden, um planungspraktische Anwendungsmöglichkeiten im Rahmen kommunaler strategischer Verkehrsplanungen zu untersuchen [Kl08]. Damit ist eine zeitlich dynamische Betrachtung von Systemveränderungen ebenso möglich wie die qualitative Wirkungsabschätzung von Maßnahmen. Daneben macht dieses Modell Wirkungsmechanismen im System transparent, was bei den gebräuchlichen Verkehrsmodellen nicht gegeben ist und immer wieder als mangelhaft beanstandet wird. Allerdings ist das Systemmodell „Stadtverkehr“ für den eigenständigen Einsatz durch Kommunalpolitiker*innen bei der konkreten Einzelfallentscheidung immer noch viel zu komplex.

Auch MobiLe wird mit Hilfe des Sensitivitätsmodells [Ve15] entwickelt. Dadurch wird eine Einbeziehung von mit dem (Verkehrs-)System verknüpften Lebensbereichen sowie eine kybernetische Analyse der resultierenden Wechselwirkungen möglich [Ve90]. Das erlaubt die – in Verkehrsmodellen bislang fehlende – qualitative Betrachtung der Wirkungen im kommunalen Kontext.

2.2 Ergebnisse aus Phase I

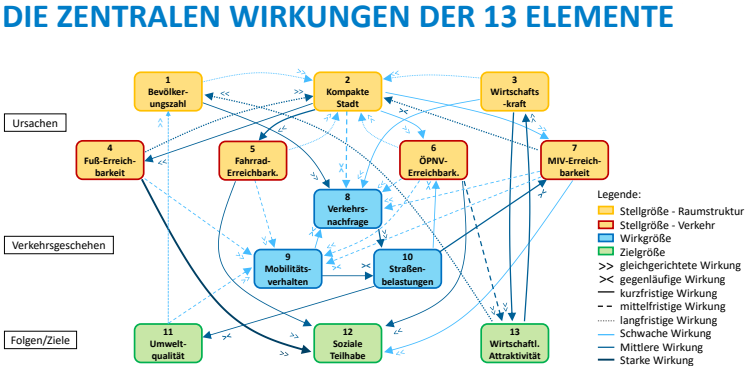


Abb. 1: Gesamtstädtisches Grundmodell für MobiLe, Stand Juni 2022

In Phase I wurde ein gesamtstädtisches Grundmodell entwickelt (Abb. 1). Es reduziert das komplexe System Verkehr auf insgesamt 13 Elemente, die sich weiter in Stell-, Wirk- und Zielgrößen unterteilen lassen. Für jedes dieser Elemente wurde geprüft, welche Einflüsse es auf andere Elemente haben könnte. Dabei wurde nicht nur beachtet, ob es Einflüsse gibt. Darüber hinaus wurden Einschätzungen vorgenommen, wie stark diese Einflüsse ausfallen, ob diese gegen- oder gleichgerichtet sind und über welchen Zeitraum der Einfluss stattfindet. Unterteilt wurde dabei in starke, mittlere und schwache Wirkung sowie in kurz-, mittel- und langfristige Einflüsse.

Bislang werden dabei ausschließlich kleinräumige Maßnahmen betrachtet, die beispielsweise innerhalb eines B-Plans Anwendung finden können oder nur eine Einzelmaßnahme wie ein neues Fahrradparkhaus darstellen. Dieser Entwicklungsschritt ist noch nicht abgeschlossen, ein Praxistest steht noch aus. Anschließend ist eine Erweiterung vorgesehen, die als zweiten Anwendungsfall großräumig wirkende Maßnahmen betrachtet, welche sich auf die gesamte Stadt auswirken – etwa ein neuer Flächennutzungsplan oder ein Radverkehrskonzept. Da es sich um ein qualitatives Modell handelt, basieren die verwendeten Einschätzungen nicht auf konkreten Zahlen, sondern spiegeln den Konsens aus der fachlichen Einschätzung verschiedener Verkehrsökologen und -wissenschaftler wider, die durch Literaturangaben so weit wie möglich gestützt wird.

Basierend auf dem Grundmodell wurde zusätzlich eine erste, noch sehr einfache Version einer Simulation entwickelt, die die Einflüsse der Elemente über einen Zeitverlauf darstellt. Eingriffe in das System können die Anwender*innen über jeweils ein Prüfschema für jede der sieben Stellgrößen vornehmen. Jedes Prüfschema enthält verschiedene fördernde und hemmende Maßnahmen für das jeweilige Element. Im Fall des Elementes Raderreichbarkeit wären dies zum Beispiel das Vorhandensein bzw. Einrichten von ausreichenden Radabstellanlagen an Quell- und Zielorten oder eine sichere Radwegeführung an Kreuzungen als fördernde Maßnahmen. Hindernde Maßnahmen oder Gegebenheiten wären unter anderem eine indirekte Wegeführung mit Umwegen oder lange Wartezeiten an (Bedarfs-)Ampeln. Für die Maßnahmen ist jeweils numerisch festgelegt, ob diese einen starken, mittleren oder schwachen Einfluss haben. Basierend auf den Summen der ausgewählten Maßnahmen erfolgt eine Umrechnung in einen starken, mittleren oder schwachen Einfluss auf den städtischen Verkehr, welcher über eine Beeinflussung der betroffenen Stellgröße anschließend in die Simulation übertragen wird. Als Ergebnis wird dargestellt, welche Auswirkungen die Maßnahmen auf die dreizehn Elemente haben. Dazu werden noch weitere Analysewerkzeuge entwickelt.

Da in Phase I aus Kostengründen mit der lizenzfreien Software Gamma 4.21 [TO] gearbeitet werden musste, sind bislang nur statische Simulationen möglich. Um dennoch einen Einblick in die Ergebnisse von dynamischen Simulationen zu bekommen, wurde behelfsweise eine einfache Programmierung in Excel vorgenommen. Damit lassen sich als Abfolge mehrerer statischer Simulationen Entwicklungen aufzeigen, die neben den direkten und indirekten Auswirkungen von Systeminterventionen auch veranschaulichen, wie sich die Zustände der Modellelemente im Zeitverlauf entwickeln. So lassen sich

qualitative Wirkungsabschätzungen von übergeordneten Strategien (z.B. einer Förderung des Radverkehrs) analysieren. Schon diese ersten einfachen Anwendungen zeigen, dass das Modell einige Auswirkungen darstellt, die nicht bereits von allen intuitiv vorhergesehen werden. Bereits das liefert einen erheblichen Mehrwert. Die unerwarteten Aussagen geben wichtige Hinweise auf bis dahin nicht erkannte Zusammenhänge und sorgen für einen Erkenntnisgewinn.

Das Modell ist zunächst auf die Stadt Norderstedt ausgelegt. Es wird für viele Städte ähnlicher Größenordnung weitgehend unverändert nutzbar sein. Um einen möglichst großen Nutzen aus dem Modell ziehen zu können, wird eine Übertragbarkeit auf andere Kommunen während der Entwicklung von Anfang an mitgeplant. Dies hat Einfluss auf die Programmierung, die dementsprechend offen entwickelt werden muss. Auf diesen Aspekt wird unten noch genauer eingegangen.

2.3 Der Prozess

Das Modell entstand in einem Co-Creation Prozess mit je zwei Vertreterinnen und Vertretern aus allen acht Fraktionen, die derzeit die Norderstedter Stadtvertretung bilden. Die dadurch erreichte Transparenz ist wesentlich für den Entwicklungsprozess und lohnt den zusätzlichen Aufwand. Ziel ist es, Vertrauen in MobiLe zu entwickeln, das die Politik gegenüber den etablierten Verkehrsmodellen gerade nicht aufbringt. Denn diese sind eine Black Box – übrigens nicht nur für Politiker*innen, sondern auch für die meisten in der Verwaltung.

Verwaltung und wissenschaftliche Partner gehen dabei stets in Vorlage. Sie erarbeiten die Grundlagen, also die Elemente des Grundmodells, die Wirkungsbeziehungen und –stärken, fördernde und hemmende Maßnahmen usw.. Das Ergebnis wird der Politik in Workshops Schritt für Schritt vorgestellt und dann miteinander diskutiert. Zum besseren Verständnis werden mitunter Beispiele aus Forschung und Verkehrsplanung vorgestellt. Jede einzelne Frage, jeder Einwand, jede zusätzliche Anregung wird bei diesem Vorgehen miteinander diskutiert und einvernehmlich geklärt. Vieles dient dabei einer weitergehenden Erläuterung der fachlichen Vorentscheidungen. Allerdings gibt es jedes Mal auch neue Erkenntnisse für die Entwickler, die zur Qualitätsverbesserung des im Aufbau befindlichen Modells beitragen. Was nicht innerhalb des Workshops zufriedenstellend beantwortet werden kann, wird von Verwaltung und wissenschaftlichen Partnern bis zum nächsten Workshop geklärt und dann wieder Punkt für Punkt erläutert und erneut diskutiert.

Es ist sehr beachtlich, dass sich die anwesenden Vertreter*innen der Fraktion hierfür die Zeit nehmen – und zwar zusätzlich zur normalen, ohnehin schon umfangreichen politischen Arbeit. Das ist jeweils ein Samstag für jeden Workshop inklusive der persönlichen Vor- und Nachbereitungen. Dieses zeitliche Engagement spricht für den Bedarf einer derartigen Unterstützung und das persönliche Interesse an guten, gut begründeten Entscheidungen. Bemerkenswert ist ferner, dass innerhalb der Workshops eine konzentrierte und kooperative

Arbeitsatmosphäre herrscht, bei der auf parteipolitische Auseinandersetzungen (bislang) völlig verzichtet wird.

Auf die obligatorische Öffentlichkeitsbeteiligung an dem Prozess ist vorerst – mit Einverständnis des BMBF – verzichtet worden. Sie wird dann folgen, wenn das Modell mit der Politik nach deren Anforderungen entwickelt und ersten Praxistests unterzogen worden ist. Dafür sprechen mehrere Gründe. Am wichtigsten erscheint, dass die Politik zunächst einen Lernvorsprung erhalten soll. Sonst ist zu erwarten, dass sie in Diskussionen auf alte Argumentationsmuster zurückgreift, die sie jahrelang eingeübt und vertreten hat. Das würde den Einsatz des Modells erheblich behindern. Die Öffentlichkeit wird sich erwartungsgemäß nur sehr kurz damit befassen, ob MobiLe für sie interessant ist. Daher muss MobiLe bereits den Status einer Betaversion erreicht haben und bis zur Webanwendung programmiert sein. Das ist noch nicht der Fall, soll aber bis 2023 gelingen. Erfahrungen aus der Einbindung der Öffentlichkeit lassen sich dann immer noch in die Optimierung von MobiLe integrieren.

3 Ziele und Anforderungen

In Phase II wird basierend auf den Entwürfen aus der vorherigen Phase zunächst die Programmierung des Grundmodells und einer dynamischen Simulation von Veränderungen des Ausgangszustands angestrebt. Das soll wieder, wie in Phase I, in einem iterativen Prozess geschehen. Die Verwaltung entwickelt die Schritte im Austausch mit den Forschungspartnern und stellt die Ergebnisse der Politik vor, mit der sie dann ausführlich diskutiert werden. Dazu gehören in zunehmendem Maße auch Anwendungstests und deren Analyse. Dabei wird das Modell dann in Einzel- und Gruppenarbeiten erprobt und entsprechend dem Feedback weiterentwickelt. Schließlich sind individuelle Anwendungen vorgesehen, deren Erfahrungen anschließend noch einmal in die Gruppe getragen, diskutiert und bei Bedarf in einer Modell-Finalisierung berücksichtigt werden.

Prozessbegleitend wird als weitere Prüf- und Testmöglichkeit ein Kommunalkreis genutzt, in dem die Übertragbarkeit und die Gültigkeit des Modells für andere Kommunen betrachtet werden. Dadurch soll die deutschlandweite Nutzbarkeit des Modells nach seiner Fertigstellung ermöglicht werden. Nicht zuletzt aus diesem Grund wird die Software als Open-Source Version angelegt. Die Methodik bei der Einbindung des Kommunalkreises orientiert sich dabei an dem Vorgehen, das sich schon in der Zusammenarbeit mit der Politik bewährt hat. Das Interesse anderer Kommunen ist groß.

Zu diesem Zeitpunkt wird es auch Testläufe mit der Öffentlichkeit geben, um auch hier zum einen die Anwendungsfreundlichkeit und zum anderen die allgemeine Akzeptanz gegenüber diesem Tool zu prüfen und zu fördern. Außerdem zielt die Einbindung der Öffentlichkeit darauf ab, die Befähigung der Bevölkerung zu einer konstruktiven Mitwirkung an der kommunalen Verkehrspolitik zu fördern. Das Ziel ist also ein Empowerment der Öffentlichkeit.

Aktuell werden bei der Entwicklung von MobiLe drei Ziele verfolgt. Das erste Ziel ist die Weiterentwicklung des bestehenden Modells. Inhaltlich wird dabei das Modell, das bisher nur auf kleinräumige Maßnahmen ausgelegt ist, auch noch auf einen gesamtstädtischen Ansatz ausgeweitet. So können auch größere Vorhaben mit ihren Wirkungen abgebildet werden, etwa ein Fußverkehrskonzept oder ein Lärmaktionsplan. Modell-technisch ist außerdem noch eine Stabilisierung der Simulation bei langfristigen Prognosen und bezüglich der Zielgröße „Soziale Teilhabe“ erforderlich. Das so weiterentwickelte Modell ermöglicht eine Einschätzung der Folgen einer Umsetzung verschiedenster Maßnahmen (Kombinationen) innerhalb Norderstedts. Dabei kann zwischen kleinräumigen und gesamtstädtischen Vorhaben unterschieden werden. Unterschiedliche Strategien können mit MobiLe schnell und vorausschauend beurteilt werden. Das kann etwa der Vergleich einer Push & Pull-Strategie mit einer gleichmäßigen Bedienung von Ansprüchen aller Verkehrsträger (anders formuliert: Wählergruppen) anhand ihrer Wirkungen sein oder die Identifizierung sinnvoller strategischer Schwerpunktsetzungen für eine Mobilitätswende mit Hilfe der Sensitivitätsanalyse. Die fachlich begründete Abschätzung der Auswirkungen von Maßnahmen erleichtert das Verständnis für die Konsequenzen innerhalb des komplexen Systems Verkehr und kann so die Entscheidungsfindung der Politiker*innen unterstützen.

Das zweite Ziel besteht in der Verbreitung von MobiLe. Durch den Co-Creation-Prozess mit Politik, Partnern aus der Wissenschaft und verschiedenen Kommunen geschieht das bereits als integrativer Teil der Modellentwicklung. Dies wird anschließend auf die Öffentlichkeit ausgeweitet. Ein Nebeneffekt davon soll sein, so früh wie möglich weitere Informationen zu sammeln und zu berücksichtigten, um die Qualität, Akzeptanz und Nutzbarkeit des Modells zu erhöhen.

Das dritte Ziel besteht in einer Programmierung von MobiLe in einer geeigneten Software und der Präsentation in einem professionellen, intuitiv eingängigen und somit anwendungsfreundlichen Design. So soll das Modell präsentiert und für die Simulation verkehrlicher Auswirkungen auf kommunaler Ebene öffentlich zugänglich gemacht werden.

4 Die Entwicklung der Software

MobiLe soll als eine dynamisierte digitale Version des Grundmodells und seiner Einstellmöglichkeiten entwickelt werden, um das kommunale Verkehrssystem zutreffend und zugleich möglichst einfach abzubilden. Die Entwicklung des Modells und der zugehörigen Software werden dabei in der agilen Entwicklungsmethodik Scrum durchgeführt. Für die Programmierung musste zunächst untersucht werden, welche Technologie sich für eine Umsetzung eignet. Erste Recherchen haben gezeigt, dass das bisher genutzte Programm GAMMA keinen öffentlichen Quellcode und keine Schnittstellen besitzt. Das erfordert die Entwicklung einer neuen Software. Daher wird die komplette Eigenentwicklung und Implementierung des Modells mit einer bekannten und anerkannten Programmiersprache angestrebt.Die Programmiersprache Python und FastAPI als Restful API werden für

die Backend-Entwicklung verwendet. Das TypeScript-basierte Angular-Webapplikations-Framework wird für das Frontend zur Visualisierung benutzt. Des Weiteren kommen Jenkins für die CI/CD-Integration und Trello für einen Scrum-basierten Ansatz zum Einsatz. Entscheidend bei der Auswahl dieser Softwaretechnologien ist zum einen, dass sie kostenlos und als Open-Source nutzbar sind und zum anderen, dass es eine große Community gibt, die die Software nutzt, wodurch die Lebenserwartung und die Qualität dieser Softwaretechnologien hoch ist. Dies ist relevant, da das Anpassen der Software für andere Kommunen mit möglichst wenig Aufwand ermöglicht werden soll, um die Übertragbarkeit der Anwendung zu erhöhen.

Die entwickelte Software kann und soll als Open-Source-Software zur Verfügung gestellt werden, um eventuelle Anpassungen für die Übertragbarkeit auf andere Städte / Kommunen zu fördern und eine weitere Nutzung auch nach Projektende zu gewährleisten.

Die Software wird als Webanwendung ausgelegt. Dadurch lässt sich die Anwendung über den Browser plattformübergreifend aufrufen, insbesondere auf Smartphones und Tablets. Dabei ist es nicht erforderlich, die Webanwendung vorab herunterzuladen oder zu installieren. Obwohl eine gute Internetverbindung erforderlich ist, benötigt die Entwicklung der Webanwendung meistens weniger Entwicklungs- und Wartungsaufwand [Ya21]. Der Ausbau der Webanwendung als Software as a Service (SaaS)-Lösung ist denkbar als ein zukünftiges Geschäftsmodell[SKP13]. Bei dieser SaaS-Lösung werden Software und IT-Infrastruktur von einem externen IT-Dienstleister betrieben und vom Kunden als Service genutzt.

Da die Software vorrangig für die Kommunalpolitik und Bevölkerung gedacht ist, muss ein besonderes Augenmerk auf die Benutzeroberfläche (User Interface, UI) gelegt werden. Nötig ist eine intuitive und responsive Gestaltung der UI. Hierzu werden die Designs und UI-Elemente von einer erfahrenen Grafikerin (MOSCHDESIGN) zugearbeitet.

Für die Entwicklung von Webanwendungen werden verschiedene Softwaretechnologien eingesetzt. Während die Datenoperationen und Berechnungen im Back-End durchgeführt werden, erfolgt die Visualisierung dazu über die Bereitstellung der Berechnungsergebnisse über das Front-End, etwa in Form von anpassbaren Graphen. Sowohl Backend als auch Frontend kommunizieren über Restful-API. Durch die Bereitstellung in Docker-basierten Containern kann die Anwendung schnell und ohne Anpassung auf verschiedene IT-Strukturen gehostet werden.

Bei der Entwicklung der neuen Software werden verschiedene Teststufen durchlaufen, nachdem ein Softwareentwurf vorliegt. So wird die Evaluation und das Testen der implementierten Software durch Unittests, Integrationstests, Systemtests und Abnahmetests von dem Nutzer über verschiedene Workshops mit den Projektpartnern durchgeführt, zu denen potenzielle Nutzer*innen eingeladen werden [SL19].

Die entwickelte Software wird den Projektpartnern bereits frühzeitig, noch während der Entwicklung zugänglich gemacht. Dazu wird das Software-Versionierungstool GitLab

verwendet, welches von den IT-Diensten der Uni Oldenburg gehostet wird. Sobald sich die Software in einer ersten lauffähigen Version befindet, wird zudem ein öffentliches GitHub Repository eröffnet und der Code für eine (Test-)Anwendung der Allgemeinheit zur Verfügung gestellt. Nach Projektabschluss ist der Code über das Repository weiterhin verfügbar. Darüber hinaus kann eine lauffähige Instanz bei der Universität Oldenburg gehostet werden. Interessierte Kommunen haben die Möglichkeit, ihre eigene Instanz mithilfe des öffentlichen Repositories aufzusetzen. Dieser Schritt kann bereits während der Projektlaufzeit geschehen und durch die Universität Oldenburg unterstützt werden. Eine Unterstützung für externe Personen bei der Bearbeitung und Anpassung des Modells nach Projektende ist denkbar. Die Art des Supports richtet sich nach dem weiteren Projektverlauf und kann daher erst zu einem späteren Zeitpunkt konkretisiert werden.

5 Ausblick

Im Verlauf der dreijährigen Projektlaufzeit wird MobiLe zu einem einfach nutzbaren Tool ausgebaut werden, das die Chance bietet, kommunale Entscheidungen im Verkehrssystem zu unterstützen und durch seinen wissensbasierten Ansatz die Abschätzung der Folgen von Entscheidungen zu erleichtern. Vor den parlamentarischen Beratungen sollen anhand der Eingabe von verkehrlich relevanten Maßnahmen durch die modellierte Vorschau schnell beurteilt werden können, welche Einflüsse diese auf das Verkehrsgeschehen in der Stadt haben. Werden die eigenen politischen Ziele damit nicht erreicht, kann über Veränderungen bei den Stellgrößen einfach nachgesteuert werden. Die Entscheidungsfreiheit der Politik wird damit nicht eingeschränkt, eine bessere Berücksichtigung der Auswirkungen kann allerdings zum besseren Verständnis der komplexen Wirkungen und zu einer Anpassung an die Erkenntnisse führen.

Der permanente Austausch mit der Politik sowie die spätere Einbindung der Öffentlichkeit führen dazu, die Akzeptanz gegenüber dem neuen Tool zu erhöhen. Darüber soll eine Nutzung auch über das Projektende hinaus gefördert werden. Die Webanwendung, die als Open-Source-Software zur Verfügung gestellt wird, kann so einen kleinen Beitrag zur notwendigen Mobilitätswende leisten.

Literatur

[Be05] Beckmann, K.; Baum, H.; Henn, A.; Klönne, M.: Szenarien und Politikstrategien für eine nachhaltige Mobilität. Abschlussbericht. - FoPS-Projekt Nr. 70.0696/2002, Aachen, Köln, 2005.

[Co20] Colville-Andersen, M.: Copenhagenize: Der Ultimative Weg zur urbanen Fahrradkultur. Thiemo Graf Verlag, 2020.

[De18] Deutscher Städtetag: Nachhaltige städtische Mobilität für alle. Agenda für eine Verkehrswende aus kommunaler Sicht. Positionspapier des Deutschen Städtetages, 2018.

[De21] Deutscher Städtetag: Monitoring zum Positionspapier „Nachhaltige städtische Mobilität für alle. Agenda für eine Verkehrswende aus kommunaler Sicht von 2018“, 2021.

[Di21] Die Bundesregierung: Deutsche Nachhaltigkeitsstrategie. Weiterentwicklung 2021, 2021.

[Ge15] Gehl, J.: Städte für Menschen. JOVIS Verlag, Berlin, 2015.

[GH20] Göllinger, T.; Harrer-Puchner, G.: 40 Jahre Neuland des Denkens: Frederic Vesters programmatische Schrift für eine nachhaltige Zukunft. IöB-Schriften, 2020.

[HWB19] Harrer-Puchner, G.; Wagener-Lohse, G.; Bobsien, A.: Co-developing and Implementing Measures for a New Mobility for Municipalities through Interconnected Thinking & System Analysis – in: Transportation Research Procedia, Vol. 41, 2019.

[Kl08] Klönne, M.: Methodik und Prozessgestaltung strategischer kommunaler Verkehrsplanungen. Anwendungsmöglichkeiten von systemdynamischen Modellen zur Bewertung kommunaler Verkehrsstrategien in der Planungspraxis, Aachen, 2008.

[Ko11] Kollosche, I.: Verkehrspolitik und Zukunftsforschung. In: Verkehrspolitik. Springer, Wiesbaden, 2011.

[Kö92] Köhl, W.; Beckmann, K.; Heberling, G.; Schmidt, R.; Becker, T.: Raumordnungspolitische Anforderungen an eine integrierte Verkehrsplanung und Verkehrsgestaltung. - Forschungsvorhaben im Auftrag des Bundesministers für Raumordnung, Bauwesen und Städtebau. Institut für Städtebau und Landes-planung der Universität Karlsruhe (TH), Karlsruhe, 1992.

[NH16] Neun, M.; Haubold, H.: The EU cycling economy. Arguments for an integrated EU cycling policy, 2016.

[SKP13] Schütz, S. W.; Kude, T.; Popp, K. M.: The impact of software-as-a-service on software ecosystems. In: International Conference of Software Business. Springer, Potsdam, 2013.

[SL19] Spillner, A.; Linz, T.: Basiswissen Softwaretest: Aus- und Weiterbildung zum Certified Tester–Foundation Level nach ISTQB®-Standard, dpunkt.verlag, Heidelberg, 2019.

[SRS02] Schade, B.; Rothengatter, W.; Schade, W.: Strategien, Maßnahmen und ökonomische Bewertung einer dauerhaft umweltgerechten Verkehrsentwicklung. Bewertung der dauerhaft umweltgerechten Verkehrsentwicklung mit dem system-dynamischen Modell ESCOT (Economic assessment of Sustainability policies Of Transport). Umweltbundesamt Berlin (Hrsg.), Berichte 5/02, Berlin, 2002.

[TO] TOPSIM: Kurzbeschreibung GAMMA – ein PC-Werkzeug für Vernetztes Denken, URL: https://www.topsim.com/gamma/.

[Ve15] Vester, F.: Die Kunst vernetzt zu denken. Ideen und Werkzeuge für einen neuen Umgang mit Komplexität. - 10. Aufl., Stuttgart, 2015.

[Ve88] Vester, F.: Hesler von, A.: Sensitivitätsmodell. 2. unveränderte Auflage, Umlandverband Frankfurt, Frankfurt, 1988.

[Ve90] Vester, F.: Ausfahrt Zukunft: Strategien für den Verkehr von morgen; eine Systemuntersuchung. Heyne, München, 1990.

[Ya21] Yaskevich, A.: Mobile web-app Im Vergleich zu anderen Arten von Apps, Sep. 2021, URL: https://www.scnsoft.de/blog/mobile-web-app-und-andere-arten-von-apps.

Digitalisierung und Innovation in der Landwirtschaft

Potentiale für innovative Lösungen auf Basis von Datenflüssen entlang von agrarischen Wertschöpfungsketten

Jantje Halberstadt [1], Arne Ortland[2], Anne-Kathrin Schwab[3], Andrii Besieda[4], Martin Kraft[5]

Abstract: Digitalisierung und Nachhaltigkeit in der Landwirtschaft befinden sich in einem komplexen Wechselspiel. Der Erfolg von digitalen Lösungen, die auf der Steigerung der Nachhaltigkeit landwirtschaftlicher Produktion abzielen, wird mittelfristig auch von der Identifikation von Innovationspotentialen entlang wertschöpfungskettenspezifischer Betriebsdatenflüsse abhängen. Die Hebung solcher Potentiale wird aktuell im Rahmen eines niedersächsischen Verbundprojektes durch inter- und transdisziplinäre Forschungsmethoden forciert. Nachfolgend werden auf der Grundlage bisher vorliegender, quantitativer und qualitativer Daten Innovationspotentiale exemplarisch für die Wertschöpfungskette der Milchviehhaltung skizziert. Anschließend wird anhand von zwei Kreativtechniken (Design Thinking und Dragon Dreaming) aufgezeigt, wie die zuvor identifizierten Innovationspotentiale mittels partizipativer Prozesse in Lösungsansätze kanalisiert werden können.

Keywords: Landwirtschaft, Digitalisierung, Nachhaltigkeit, Niedersachsen, Wertschöpfungsketten, Betriebsdatenflüsse, geschützte Transparenz, Transaktionskosten

1 Einleitung

Trotz zahlreicher Vorteile gehen mit Digitalisierung in der Landwirtschaft auch Herausforderungen einher, denen mittelfristig begegnet werden muss. Mit der Implementierung neuartiger, digitaler Lösungen, wie etwa Farm-Management- und -Informationssysteme (FMIS) werden einerseits die technische Durchlässigkeit der beteiligten Wertschöpfungsnetze für die anfallenden Daten gefordert. Andererseits haben aber die beteiligten Akteure ein berechtigtes Interesse an der Wahrung ihrer Datenhoheit. Der hier entstehende Konflikt zwischen Datentransparenz und Datensicherheit resultiert in der Forderung nach „ge-

[1] Universität Vechta, Fakultät I Wirtschaft und Ethik, Professur für Ökonomie der Nachhaltigkeit, Driverstraße 22, 49377 Vechta, jantje.halberstadt@uni-vechta.de
[2] Universität Vechta, Fakultät I Wirtschaft und Ethik, Driverstraße 22, 49377 Vechta, arne.ortland@uni-vechta.de
[3] Universität Vechta, Fakultät I Wirtschaft und Ethik, Driverstraße 22, 49377 Vechta, anne-kathrin.schwab@uni-vechta.de
[4] Thünen-Institut für Agrartechnologie, Bundesallee 47, 38116 Braunschweig, andrii.besieda@thuenen.de
[5] Thünen-Institut für Agrartechnologie, Bundesallee 47, 38116 Braunschweig, martin.kraft@thuenen.de

doi:10.18420/inf2022_68

schützter Transparenz“. Viele der verfügbaren Lösungen bauen auf proprietären Softwaresystemen auf und es können somit Kakophonie-Effekte entstehen, die die Qualität der Datennutzung weiter mindern. Gleichzeitig wächst der externe Druck auf die Landwirtschaft, in einem stark marktwirtschaftlich orientierten System nachhaltig zu produzieren. Bisher werden hier die Potentiale der Digitalisierung, zum Beispiel in Form einer automatisierten Weitergabe von Tierwohl- und Cross-Compliance-relevanten Daten, aufgrund diverser Faktoren nicht voll ausgenutzt.

Diese Herausforderungen sollten im Sinne unternehmerischen Denkens und Handelns als Chancen für die Ableitung innovativer Lösungsansätze gesehen werden. Das bildet den Ausgangspunkt eines inter- und transdisziplinären Forschungsprojekts, dessen Fokus auf der ganzheitlichen Betrachtung von Digitalisierungspotentialen in der niedersächsischen Landwirtschaft liegt. Speziell für Niedersachsen gilt, dass große Regionen stark agrarisch geprägt sind (mit ökonomischen wie ökologischen und sozialen Konsequenzen). Für diese Regionen werden in einem Teilprojekt sämtliche Datenflüsse der involvierten Stakeholder erhoben sowie mögliche Lösungsansätze diskutiert und auf ihre Übertragbarkeit zu anderen Regionen untersucht.

Im Rahmen der Konferenz soll das Forschungsprojekt vorgestellt und erste Ergebnisse präsentiert werden. Anhand der Wertschöpfungskette der Milchwirtschaft wird aufgezeigt, wie sich Potenziale für innovative Lösungen ableiten lassen. Schließlich werden zwei partizipative Kreativtechniken vorgestellt, die sich dazu eignen, die identifizierten Potentiale unter Beteiligung wichtiger Stakeholder in innovative Lösungsideen umzuwandeln.

2 Digitalisierung in der Landwirtschaft: Chancen und Herausforderungen

Digitale Technologien finden zunehmend Eingang in die landwirtschaftliche Produktion. Mit der Einführung von Computern und Sensoren in die Landwirtschaft um 1980 (Schön, 1993), hat sich diese rasant bis zum derzeitigen „Precision Farming“ und „Precision Livestock Farming“ weiterentwickelt. Eine Vielzahl unterschiedlichster Sensoren erheben kontinuierlich Daten, die auf verschiedenen Plattformen gespeichert und mit unterschiedlichen Softwaresystemen verarbeitet werden (Stiene et al., 2017). Neben den innerbetrieblichen Sensorsystemen in der Nutztierhaltung und im Pflanzenbau, die primär der nachhaltigen Betriebsoptimierung dienen, sind landwirtschaftliche Betriebe gesetzlich zur Aufzeichnung und Weiterleitung einer Vielzahl von Daten an staatliche Institutionen und Verbänden verpflichtet (Haase, Kluge, 2017).

Laut einer deutschlandweiten Umfrage (n=591) von Landwirt*innen durch die Bayerische Landesanstalt für Landwirtschaft (LfL) am Jahresende 2019 nutzt etwa ein Drittel der Landwirt*innen Farm Management Information Software (FMIS) (Gabriel et al., 2021).

Automatische Systeme haben in der Außenwirtschaft (z.B. Lenksysteme, Teilbreitenschaltung) einen Verbreitungsgrad von teilweise über 30 %. In der Innenwirtschaft haben beispielsweise automatische Melksysteme und voll autonome Spaltenreinigungsroboter einen Verbreitungsgrad von über 15 %. Technologien, die dem Bereich des Smart Farming oder Precision Farming zuzuordnen sind, sind ebenfalls auf dem Vormarsch. Beispielsweise gaben 24 % der Umfrageteilnehmer*innen aus dem Bereich Außenwirtschaft an, dass sie Karten aus Satellitendaten nutzen. In der Innenwirtschaft nutzen 20 % der Umfrageteilnehmer*innen bereits Sensoren zur Verhaltensüberwachung.

Es gibt jedoch auch einige Herausforderungen. Wesentliche Hemmnisse für die Einführung digitaler Technologien in der Landwirtschaft sind aus Sicht der Landwirt*innen der hohe Investitionsbedarf, die fragliche Wirtschaftlichkeit und die umständliche Bedienung. Ein häufig artikuliertes Problem stellt die Vernetzung von Systemen und Prozessen dar. Die Teilnehmer*innen der o.g. Umfrage brachten zum Beispiel zum Ausdruck, dass die Systeme verschiedener Hersteller nicht interoperabel sind und die Vernetzung zusätzlich durch die mangelhafte Versorgung des ländlichen Raums mit mobilen Daten erschwert wird (VDI, 2021). Das Datenmanagement der umfangreich zur Verfügung stehenden, aber nicht einheitlich erhobenen und bearbeitbaren Daten birgt entsprechend zahlreicher Herausforderungen (European Union 2015; Martínez, 2016). Schon jetzt ist u.a. dadurch der bürokratische Aufwand für die Landwirtschaft und auch für die öffentliche Verwaltung als hoch einzuschätzen (El Benni et al., 2021; Latacz-Lohmann & Schreiner, 2017). Verstärkt wird dieser Effekt aktuell durch die Nutzung heterogener und zum Teil inkompatibler Datenmanagement-Lösungen ohne gängige Standards (Scheuren et al., 2021; Heinze et al., 2014).

Mehr und mehr beeinflussen zudem ungeklärte Fragen in Bezug auf Datenhoheit und Datensicherheit die Entscheidung zur Anschaffung und Nutzung digitaler Systeme (Gabriel et al., 2021; Krüger, 2016; Landschof, 2020). Dies gilt auch und ganz besonders für Digitalisierung in der Landwirtschaft (Schmidt, 2018; Umstätter et al., 2020). Viele Prozesse in der Agrarwirtschaft lassen sich nur durchführen, wenn Daten den Betrieb verlassen und zwischen verschiedenen Akteuren der Wertschöpfungsketten ausgetauscht werden (Transparenz). Andererseits besteht ein berechtigtes Interesse daran, dass die Datenhoheit bei den Urheber*innen bleibt und nicht ungewollt Informationen weitergegeben werden. Wie Datenfluss gewährleitet und trotzdem Datenhoheit durch die Landwirt*innen über sensible Daten realisiert werden kann, wird unter dem Begriff der „geschützten Transparenz“ diskutiert und ist bereits ein etabliertes Forschungsthema (Bitkom, 2016; BMEL 2017; Hertzberg et al., 2020).

Um die konkreten Probleme im Zusammenhang mit Datenflüssen in der Landwirtschaft besser zu verstehen und Lösungen dafür zu entwickeln, sollten Datenströme entlang vollständiger Wertschöpfungsketten in den Blick genommen werden. Mit einer gesamtheitlichen Betrachtungsweise können Defizite gesetzlicher und freiwilliger, sowohl interner als auch externer Datenflüsse aufgezeigt werden. Dadurch entstehen nicht nur neue Ansätze zur Optimierung des gesetzlichen Meldewesens sowie der praktischen Realisierung eines

„Smart Farming“, sondern optimalerweise auch konkrete Vorschläge für praktische Verbesserungen. Dazu wird im Folgenden anhand der Milchwirtschaft als Beispielwertschöpfungskette kurz skizziert, welche Innovationspotentiale sich aus dem derzeitigen

3 Wo Daten entstehen und ausgetauscht werden: Potenziale identifizieren anhand von Wertschöpfungskettenanalysen

Um Datenflüsse zu identifizieren und zu analysieren, wurde ein Mixed-Methods-Ansatz gewählt, der quantitative und qualitative Methoden miteinander kombiniert (Baur et al., 2017). Ausgangspunkt stellten Wertschöpfungsketten verschiedener landwirtschaftlicher Produktionsformen (Milchvieh-, Schweine- und Hühnermastbetriebe) und die dort identifizierten wesentlichen Stakeholder und Verarbeitungsschritte dar. Ergänzt wurden dies um eine Online-Umfrage zur quantitativen Analyse des Automatisierungsgrades verschiedener Betriebsdatenflüsse und gesetzlicher Dokumentationspflichten sowie der sich hieraus ableitenden zeitlichen und monetären Kosten. Aufbauend auf den Erkenntnissen der Online-Umfrage wurden leitfadengestützte Experteninterviews konzipiert (Witzel, 2000). Durch diesen Ansatz soll ergänzend die Wiedergabe von Meinungen und Erfahrungen in den Vordergrund gerückt werden (Kramer, Pfaffenbach, 2018). Die Auswertung der Interviews orientiert sich an den Grundlagen der qualitativen Inhaltsanalyse (Mayring, Fenzl, 2019). Exemplarisch werden erste Ergebnisse der Analyse der Datenflüsse in den Milchviehbetrieben vorgestellt.

Die Milchwirtschaft zeichnet sich durch eine Abfolge komplexer Arbeitsschritte aus, die jeweils von unterschiedlichen technisch-organisationalen Faktoren geprägt sind. Ähnlich wie auf Mastbetrieben wird auch in der Milchwirtschaft jeder Wertschöpfungsschritt eng von privaten und staatlichen Stakeholdern begleitet. Der Wertschöpfungsprozess wird darüber in seiner Gesamtheit von staatlichen Prüfstellen kontrolliert sowie von Zucht- und Berufsverbänden sowie landwirtschaftlichen Beratungsstellen begleitet (siehe Abb. 1).

Die Zusammenarbeit mit entsprechenden Stakeholdern ist wiederum eng gekoppelt an die Erfassung, Speicherung und Weitergabe von Betriebsdaten. Es wurde im Rahmen der Online-Umfrage von den teilnehmenden Milchviehbetrieben die Weitergabe von Milchleistungs- und Milchgütedaten, insbesondere an Molkereien, stark betont. Darüber hinaus spielt neben universellen Datenpositionen, wie etwa administrativen Betriebsdaten oder Daten der Betriebsausstattung für Milchviehbetriebe insbesondere die Weitergabe von Vital- und Besamungsdaten an Bestandstierärzte eine zentrale Rolle.

Bei der Betrachtung des Digitalisierungs- beziehungsweise Automatisierungsgrades des wertschöpfungskettenspezifischen Austauschs von Betriebsdaten mit Stakeholdern in der Milchviehhaltung fällt auf, dass insbesondere die Erfassung von tierbezogenen Daten (Zuchtwerte, Milchleistungs- und Milchgütedaten, Vitaldaten) bereits einem hohen Automatisierungsgrad unterliegt. Hierbei handelt es sich um solche Betriebs- und Tierdaten,

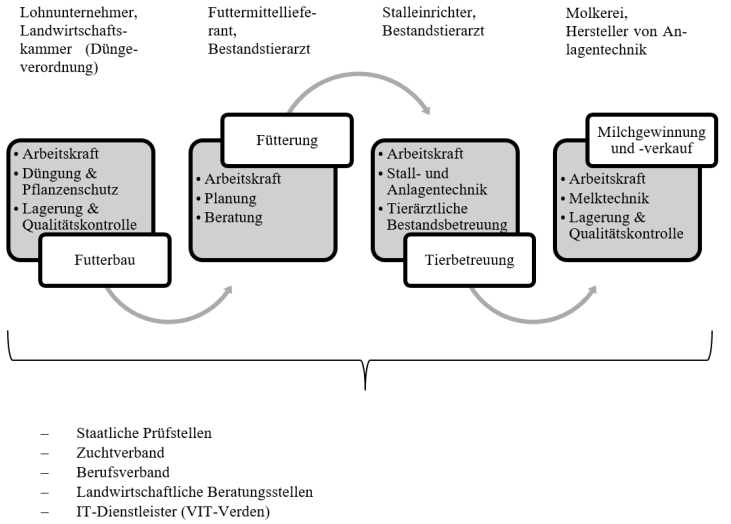
deren Erfassung durch das Herkunftssicherungs- und Informationssystem für Tiere (abgekürzt HI-Tier oder HIT) oder aber durch digitale Plattformen der Molkereien standardisiert geregelt wird. Demgegenüber fällt auf, dass die Erfassung und Weitergabe von Behandlungsdaten auf vielen Betrieben noch überwiegend händisch zu erfolgen scheint.

Teils fehlende Möglichkeiten einer automatisierten Weitergabe von Cross-Compliance-relevanten Daten hat insgesamt zur Folge, dass der Datenaustausch mit staatlichen Stellen (z.B. Veterinäramt) häufig überwiegend manuell erfolgt und zudem von Medienbrüchen geprägt ist, die das Datenmanagement für Landwirt*innen zusätzlich erschweren. Trotz dieses Defizits scheint der Austausch von Cross-Compliance-relevanten Betriebs-daten mit staatlichen Stellen hingegen für die Mehrzahl der derzeit verfügbaren Software-Lösungen aktuell keine nennenswerte Relevanz zu spielen. Es muss angenommen werden, dass die oftmals fehlende digitale Verzahnung von Betrieb und Prüfstelle den bürokratischen Zeitaufwand für Landwirt*innen signifikant erhöht (El Benni et al., 2021).

Bislang setzen viele Software-Lösungen, wie etwa der Agrirouter, ihren Schwerpunkt auf die Erhebung und Übermittlung telemetrischer Daten (z.B. Applikationsdaten, Ernte- und Ertragsdaten, Prozessdaten, Bodendaten). Das hierbei angestrebte Ziel sind demnach Produktivitäts- und Effizienzsteigerungen durch die herstellerübergreifende Vernetzung von Landmaschinen und Agrarsoftware (Homepage Agrirouter 2022). Auch die anbieterunabhängige Software-Lösung 365FarmNet setzt vor allem auf die digitale Vernetzung von Betriebsdaten zur Optimierung des Betriebsmanagements, des Pflanzenbaus und des Flotten- und Herdenmanagements (Homepage 365FarmNet, 2022). Obwohl sich hieraus zweifelsohne positive Implikationen für Naturschutz und Tierwohl ergeben, spielt auch hier das Management Cross-Compliance-relevanter Daten hingegen aktuell noch keine nennenswerte Rolle.

Es zeigt sich in den bisherigen Analysen ein komplexes Spannungsfeld, in dem die geschützte Transparenz angesiedelt ist: Transparenz und Datendurchlässigkeit stehen Datenschutz und Datenhoheit gegenüber. Eine ‚smarte' Kommunikation von Cross-Compliance-relevanten Daten, die zur Reduktion der Transaktionskosten beiträgt, kann nur erzielt werden, wenn die technischen Möglichkeiten der Überprüfung und Nachverfolgung von staatlicher Seite nicht zu einer umfassenden Kontrolle und Überwachung führen.

Ein Beispiel dafür, wie ein solcher Datenaustausch gelingen kann, wäre zum Beispiel das Projekt ODiL (Offene Software-Plattform für Dienstleistungsinnovationen in einem Wertschöpfungsnetz in der Landwirtschaft), das mit Mitteln des Bundesministeriums für Bildung und Forschung (BMBF) im Zeitraum 2016-2019 gefördert wurde. Hier wurde eine dezentrale Datenplattform mit bedarfsorientierter Öffnung von gesicherten Datenkanälen entwickelt (Stiene et al., 2017), welche heute als Prototyp zur Verfügung steht (Homepage ODiL, 2020). Demnach ist die geringe Automatisierung der Datenübermittlung im Kontext gesetzlicher Dokumentationspflichten nicht primär das Resultat einer technischen Komplexität. Vielmehr gilt es, verstärkt Akzeptanzfragen in die Diskussion einzubeziehen. Wie dies gelingen kann, zeigt das nachfolgende Kapitel.



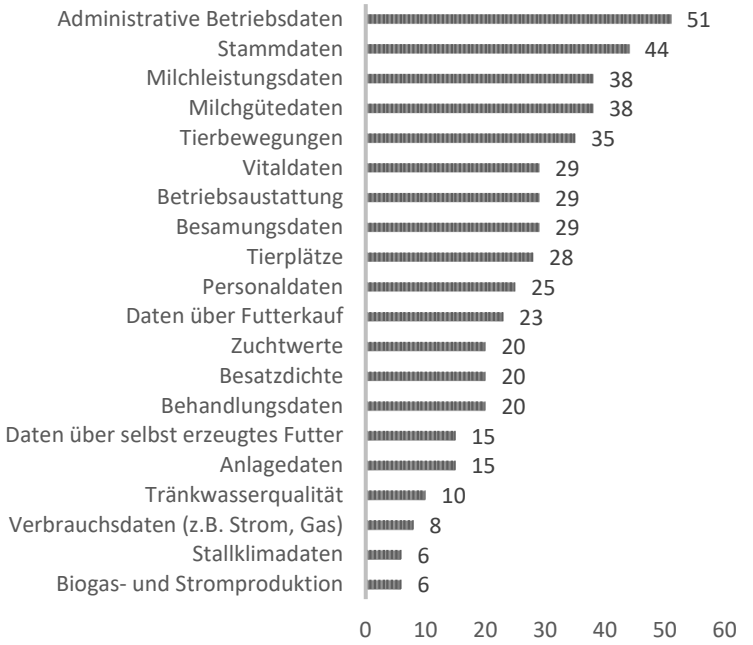


Abb. 1: Wertschöpfungskette der Milchwirtschaft: beteiligte Stakeholder und geteilte Daten (n=10), Quelle: eigene Daten, Erklärung: das untenstehende Diagramm nennt die absoluten Häufigkeiten, mit der Stakeholder in der Online-Umfrage als relevant für Betriebsdatenflüsse identifiziert wurden; es konnten pro Datenkategorie bis zu 10 Stakeholder ausgewählt werden.

4 Wie Innovationspotenziale gehoben werden können: Zwei Ansätze zur partizipativen Generierung von nachhaltigen Lösungsideen

Um die identifizierten Potentiale in vollem Ausmaß nutzbar machen zu können, müssen Lösungen generiert, diskutiert, bewertet und schließlich umgesetzt werden. Die Generierung von Lösungsideen stellt die Basis für alle weiteren, darauf aufbauenden Aktivitäten dar. Von zentraler Bedeutung hierbei ist es, Ideen zu generieren, die bereits die zentralen Probleme adressieren und gleichzeitig die jeweiligen Rahmenbedingungen mitdenken. Das bedeutet, dass man vor der Herausforderung steht, einen Rahmen für kreativen Freiraum zu schaffen, welcher gleichzeitig mögliche Restriktionen, etwa in der Form knapper Ressourcen oder rechtlicher Vorgaben, berücksichtigt.

Eine Möglichkeit liegt in partizipativen Prozessen (Von Unger, 2014). Das bedeutet, wichtige Stakeholder, hier z.B. Landwirt*innen, Wirtschaftspartner*innen und Vertreter*innen staatlicher Stellen in Kreativprozesse einzubinden. Für entsprechende Workshop-Formate kommen dann verschiedene Wege der Ausgestaltung in Frage. Im Folgenden möchten wir zwei durchaus unterschiedliche Methoden vorstellen, die sich dazu eignen, kreative und partizipative Elemente miteinander zu verknüpfen: Design Thinking und Dragon Dreaming.

4.1 Design Thinking

„Design ist nicht nur, wie es aussieht und sich anfühlt. Design ist, wie es funktioniert."

-Steve Jobs

Beim Design Thinking geht es darum, Methoden, die von Designer*innen angewandt werden, für Ideengenerierungsprozesse zu nutzen, um so neue Problemlösungsperspektiven zu erzeugen. Die Kreativtechnik baut dabei auf drei Grundprinzipien auf: Kreativität, Multiperspektivität, Prozesshaftigkeit. Heruntergebrochen bedeutet dies, dass Probleme in einem kreativen Setting aus möglichst verschiedenen Perspektiven heraus erforscht und in einem wiederkehrenden (iterativen) Prozess mögliche Lösungsideen immer wieder kritisch hinterfragt werden (Freiling et al., 2019; Schawel, Billing, 2014).

Aus einem grundsätzlich abduktiven Denken heraus ist Design Thinking empathisch, persönlich, subjektiv, interpretativ, integrativ, experimentell, synthetisch, bildhaft und dialektisch (Serrat, 2017). Design Thinking besteht aus drei Schlüsselphasen: Inspiration, Ideenfindung und Umsetzung. Je nach wissenschaftlicher Disziplin, in der Design Thinking implementiert wird, lassen sich auch sechs oder sieben Phasen mit teils unterschiedlicher Betitelung unterteilen. In diesen verschiedenen Phasen werden Probleme formuliert, Fragen gestellt, Ideen entwickelt und Antworten gefunden. Die Phasen verlaufen nicht linear, sondern können gleichzeitig ablaufen oder wiederholt werden (Martin, 2009).

Design Thinking ist gleichzeitig von Natur aus eine Art Prototyping-Prozess, der ein möglichst tiefes Verständnis darüber aufbauen möchte, welche Ansprüche potentielle Nutzer*innen auf eine Lösung projizieren und warum sie dies tun (siehe Abbildung 2). Es wird davon ausgegangen, dass Innovationen nicht durch inkrementelle Optimierungen entstehen und ihrem Wesen nach auch nie abgeschlossen sind. Marktveränderungen, technologischer Fortschritt, gesetzliche Rahmenbedingungen, Umweltdynamiken und viele weitere Faktoren führen dazu, dass die Effektivität vermeintlich innovativer Lösungsansätze nicht immer vorherzusehen ist und mitunter gar neue Probleme aufwirft. Nur ein solches Grundverständnis erlaubt mittelfristig die Entwicklung von Lösungen zu komplexen, multifaktoriellen Problemen (Van Berkel, Manickam, 2020; Serrat, 2017).

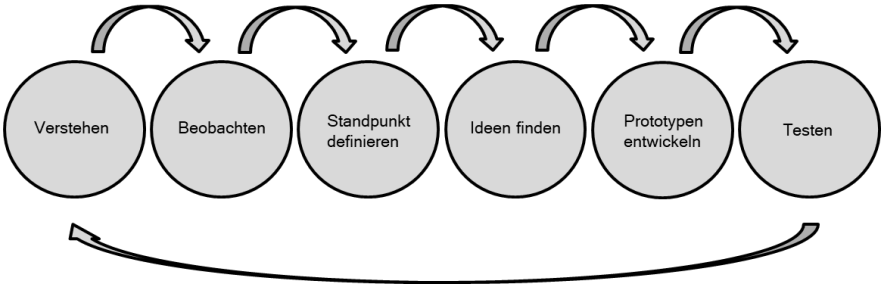


Abbildung 2: Phasen des Design Thinking, eigene Darstellung nach Freiling et al. 2019; Schawel, Billing 2014

4.2 Dragon Dreaming

Ähnlich wie beim Design-Thinking-Prozess handelt es sich beim Dragon Dreaming um eine Projektarbeitsmethode, die ebenfalls verschiedene Perspektiven im Problemlösungsprozess forciert (siehe Abbildung 3). Die intellektuellen Kernelemente des Dragon Dreaming sind den indigenen Lebensweisheiten der westaustralischen Aborigines entlehnt und valorisieren Selbstermächtigung und Gemeinschaftsbildung, Nachhaltigkeit, Mobilisierung und eine innovative Projektentwicklung (Croft, 2011). Der Dragon-Dreaming-Prozess basiert auf einer weitgefassten Auffassung davon, was ein Projekt sein kann. Kern des Prozesses ist allerdings nicht die Methode, sondern die Individuen, die an diesem teilhaben: Es geht nicht nur um das Resultat, sondern auch um die Vernetzung und persönliche Weiterentwicklung der individuellen Teilnehmer*innen. Theoretisches Wissen und praktische Fertigkeiten spielen eine gleichberechtigte Rolle. Die folgenden vier Phasen prägen den Projektkreislauf und werden allgemeinhin auch als „Traumkreis“ bezeichnet (Croft, 2011):

1. Träumen – Informationen sammeln, Motivation klären, Bewusstsein schaffen
2. Planen – Alternativen abwägen, Strategien entwickeln, Testen und Prototypen
3. Handeln – Implementierung, Administration und Management, Monitoring
4. Feiern – Ergebnisse transformieren, neue Fähigkeiten und Weisheit erlangen

Das Innovative an dieser Methode ist, dass das „Feiern“ hierbei im Kreislauf einen gleichberechtigten Platz neben all den anderen Phasen einnimmt. Es kann eine reflexive und wertschätzende Qualität haben, aber auch ein Bewusstsein darüber, welche neuen Fähigkeiten, welches Wissen während des Projektprozesses erlangt wurden. Der fraktale Kreislauf im Dragon-Dreaming-Prozess bedeutet, dass in jedem der vier Projektphasen alle Phasen stattfinden können, das heißt auch im Träumen findet Planen, Handeln und Feiern statt: das Vorgehen kann daher auch als spiralförmig und nichtlinear bezeichnet werden.

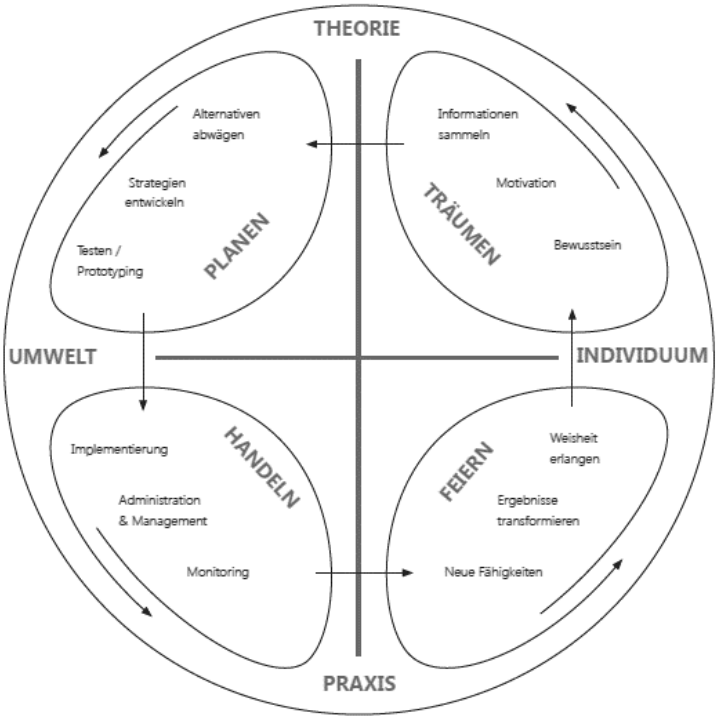


Abb. 3: Der Projektkreislauf des Dragon Dreaming nach Koglin (2012) und Croft (2011)

Im Kontext des Projektes kann das Träumen zum Beispiel die Idealvorstellungen von Datenerhebung und Datenübermittlung aus der Sicht der verschiedenen Stakeholder enthalten. Das Besondere ist, dass alle Phasen zeitlich die gleiche Aufmerksamkeit erhalten. Ebenso ist es mit den Teilnehmer*innen des Projektes: Auch hier geht es darum, Träumer*innen, Planer*innen, Macher*innen und Feierer*innen die gleiche Anerkennung zu schenken und alle Typen als wertvolle, gegensätzliche Funktionen innerhalb des Projektes und des Prozesses wahrzunehmen (Croft, 2011).

Im Dragon Dreaming Prozess spielen unter anderem Übungen der Stille, aber auch der „charismatischen Kommunikation", also einer Kommunikation, die aus einem Gefühl heraus geschieht, eine wichtige Rolle. Dies kann insbesondere bei sensiblen Themen, wie etwa der Digitalisierung von Betriebsdatenflüssen von Vorteil sein, um in einer vermeintlich sachlich-zielorientierten Diskussion Emotionalität zu abstrahieren. Eine relevante Einstiegsfrage zu einem Traumkreis, etwa im Kontext der Digitalisierung von Cross-Compliance-relevanten Betriebsdatenflüssen (siehe Kapitel 3) könnte daher lauten:

„Wie kann die automatisierte Erfassung und Weitergabe von Betriebsdaten dem Anspruch einer Datenhoheit durch den landwirtschaftlichen Betrieb wahren und dabei gleichzeitig eine größtmögliche Transparenz gegenüber staatlichen Prüfstellen und Wirtschaftspartnern ermöglichen?"

Die Art und Weise, in der die obige Frage formuliert ist, setzt voraus, dass möglichst alle Stakeholder entlang der Wertschöpfungskette Teil des kreativen Denkprozesses sind. Die Aufgabenplanung ist dabei in der Form eines Spinnennetzes organisiert. Es werden Zeit- und Kostenpläne erstellt. Alle Aufgaben und Verantwortlichkeiten, Zeit, Budget werden in dem sogenannten Projekt-Status-Barometer zusammengetragen und geplant. Die Aufgaben werden durch sogenannte „Slonglines" miteinander verbunden. „Slonglines" sind ein wichtiger Indikator für Ausgewogenheit und Nachhaltigkeit innerhalb der Projektplanung. In der Aufgabenverteilung der jeweiligen Projektpakete gibt es Projektverantwortliche, Trainees und Berater, die durch einen gemeinsamen Prozess im Hinblick auf die Aufgaben festgelegt werden (vgl. Koglin, 2012).

Wenn die Dragon-Dreaming-Technik für Innovationsworkshops zum Themen wie „Optimierung von Datenflüssen in der Landwirtschaft" oder „Verbesserte Nutzung von Daten in der Landwirtschaft" angewandt werden soll, bedarf es einer zielgerichteten und vor allem auf die Teilnehmenden angepassten Adaption. Die Phasen des Handelns und des Feierns werden könnten beispielsweise in einer ersten Sitzung nur gestreift werden, wobei der Rahmen einen „feierlichen" Austausch zwischen den verschiedenen Personen ermöglichet und betont werden soll. Es ist zudem wichtig, die Teilnehmenden zu sensibilisieren und in die Technik einzuführen.

5 Zusammenfassung und Ausblick

In der Digitalisierung in der landwirtschaftlichen Produktion liegen zahlreiche Potenziale. Beispielsweise kann eine automatisierte Erfassung und Weitergabe von Tierwohl-relevanten Cross-Compliance-Daten einen positiven Beitrag zum eingeforderten Nachhaltigkeitsdiskurs leisten und gleichzeitig die Transaktionskosten für Landwirt*innen reduzieren. Von einer technischen Integration dieses Ansatzes in bestehende Software-Lösungen (Agrirouter oder 365FarmNet), können darüber hinaus auch weitere Stakeholder in Produktion, Handel und Verarbeitung profitieren (Gandorfer et al., 2017). Gleichzeitig werden zahlreiche Herausforderungen identifiziert, vor denen Landwirt*innen und andere Stake-

holder stehen. Diese reichen von inkompatiblen Systemen bis hin zu Datensicherheitsproblemen. Im Sinne unternehmerischen Denkens und Handelns sollten diese Probleme jedoch als Chancen gesehen werden, um innovative Lösungen abzuleiten. Zudem liegt weiteres ungenutztes Potenzial etwa in der Vielzahl von Daten, die bislang nicht oder nicht ausreichend aufgearbeitet und genutzt werden.

Auf der Konferenz soll ein Forschungsprojekt vorgestellt werden, das diese Potenziale in den Blick nimmt. Erste Ergebnisse der Analyse von Datenflüssen werden am Beispiel der Milchwirtschaft vorgestellt. Entlang der Wertschöpfungskette wurden zunächst diverse Daten triangulativ erfasst und deskriptiv dargelegt. Nun folgen Innovations-Workshops, mithilfe derer innovative Lösungen abduktiv und partizipativ generiert werden. Theorie und Praxis werden dabei gleichermaßen einbezogen und innerhalb eines kreativ-partizipativen Prozesses adressiert. Akzeptanzhemmnisse der Stakeholder oder Interoperabilität an Knoten des Wertschöpfungsnetzes, sowie Schnittstellenprobleme oder Medienbrüche können so multiperspektivisch adressiert werden. Geteiltes Interesse aller Stakeholder muss es dabei sein, „analoge Absurditäten“, wie etwa das Ausdrucken digitaler Daten, um diese postalisch oder per Fax zu übermitteln, zu vermeiden (Hirt und Bröckl, 2020).

Durch einen partizipativen Austausch innerhalb eines innovativen Formates werden auch emotional sensible Themen, wie die Veränderung von Arbeitsplätzen oder die durch die Digitalisierung verursachten Umbrüche von Betriebsstrukturen, adressiert. Ebenso kann die gesellschaftlich formulierte Kritik an der landwirtschaftlichen Produktion mit in die Debatte aufgenommen werden. Der Einbezug von Praktiker*innen der verschiedenen Schnittstellen und Knotenpunkte im Wertschöpfungsprozess ermöglicht eine Erfassung der Komplexität der Digitalisierung in der Landwirtschaft (Nüssel, 2018). Die Adressierung der technischen, aber auch sozialen Fragen an einen heterogenen Teilnehmer*innenkreis von Landwirt*innen, Maschinenhersteller*innen, Dienstleistenden, der Lebensmittelindustrie und der Konsument*innen ermöglicht das Erkennen und die Adressierung bestehender und zukünftiger Konfliktfelder. Schlussendlich kann eine Entzerrung dieser Konfliktfelder nur dann gelingen, wenn ein gemeinschaftlicher, sektorenübergreifender Blick auf die Ursachen forciert wird (Fuchs et al., 2020; Van Berkel, Manickam, 2020). Daher werden auf der Konferenz zwei Möglichkeiten des Einsatzes von Kreativtechniken vorgestellt und diskutiert, mit Hilfe derer Rechnung getragen werden kann.

Quellenverzeichnis

Agrirouter, https://my-agrirouter.com/de/agrirouter/grundbegriffe/, Stand 13.06.2022.

Bahrs, E.: Exemplarische betriebswirtschaftliche Auswirkungen der Digitalisierung in der Landwirtschaft und im Agribusiness. Wilhelm-Schaumann-Stiftung, 2018.

Bardmann, M.: Grundlagen der Allgemeinen Betriebswirtschaftslehre. Geschichte – Konzepte – Digitalisierung. 3., vollständig überarbeitete und erweiterte Auflage. Wiesbaden: Springer Fachmedien, 2019.

Baur, N.; Kelle, U.; Kuckartz, U.: Mixed Methods. KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie. Edition: Sonderheft 57, 481, Springer, 2017.

Bitkom, http://www.bitkom.org/noindex/Publikationen/2016/Positionspapiere/Digitalisierung-in-der- Landwirtschaft/Bitkom-Positionspapier-Digitalisierung-in-der-Landwirtschaft.pdf, Stand: 14.06.2022.

BMEL, Digitalpolitik Landwirtschaft. Zukunftsprogramm: Chancen nutzen – Risiken minimieren. www.bmel.de/SharedDocs/Downloads/Broschueren/DigitalpolitikLandwirtschaft.html, Stand: 14.06.2022.

European Union. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52015SC0100, Stand 14.06.2022.

Croft, J.:Dragon Dreaming: Eine partizipative Methode für die Entwicklung gemeinschaftsorientierter nachhaltiger Projekte. Zeitschrift Bewusstseinswissenschaften. Transpersonale Psychologie und Psychotherapie, 17(1), 69-79, 2011.

Dilger, T.; Ploder, C., & Bernsteiner, R.: Smart Farming - Herausforderungen und Potenziale in der kleinstrukturierten Landwirtschaft Österreichs. Responsible Technology - Nachhaltige Technologieansätze für eine verantwortungsvolle Gestaltung der Zukunft, 2021

El Benni, N.; Ritzel, C.; Heitkämper, K.; Umstätter, C.; Zorn, A.; Mack, G.: The cost of farmers' administrative burdens due to cross-compliance obligations. Journal of Environmental Planning and Management 65(5), 930-952, 2021.

Freiling, J.; Harima, J.: Entrepreneurship – Gründung und Skalierung von Startups, Springer Fachmedien, Wiesbaden, 531, 2019.

Fuchs, C.; Gütschow, P.; Langosch, R.; Skau, K.; Thierbächer, L.: Verbesserung der Klimabilanzen landwirtschaftlicher Betriebe und der Beitrag der Digitalisierung. HMD Praxis für Wirtschaftsinformatik, 58, 83-101, 2020.

Gabriel, A.; Gandorfer, M.; Spykman, O.: Nutzung und Hemmnisse digitaler Technologien in der Landwirtschaft: Sichtweisen aus der Praxis und in den Fachmedien. Berichte über Landwirtschaft, 99 (1), 2021.

Gandorfer, M.; Schleicher, S.; Heuser, S.; Pfeiffer, J.; Demmel, M.: Landwirtschaft 4.0– Digitalisierung und ihre Herausforderungen. Ackerbau-technische Lösungen für die Zukunft 6, 9-21, 2017.

Haase, M. S.; Kluge, V.: Rechtliche Bewertung der zunehmenden Informationsverarbeitung in der digitalisierten Landwirtschaft. Gesellschaft für Informatik e.V., Bonn, 65-68, 2017.

Heinze, C.; Haidn, B.; Bernhardt, H.: Nutzung von Herdenmanagement-Software in der Milchvieh-haltung. In 19. Arbeitswissenschaftliches Kolloquium des VDI-MEG Arbeitskreises Arbeitswissenschaften im Landbau, 116, 2014.

Hertzberg, J.; Heijne, D.; Kisliuk, B.; Scheuren, S.; Stiene, S.: Geschützte Transparenz. Über technische und andere Voraussetzungen für die Digitalisierung der Landwirtschaft. In Digitalisierung, Automatisierung, KI und Recht, 59-80, 2020.

Hirt, M.; Bröckl, A.: Man muss sich auf das System verlassen können, sonst gibt es Radau im Stall. AMS info No. 462, 5, 2020.

Homepage 365 FarmNet. https://www.365farmnet.com/de/, Stand 28.06.2022.

Homepage ODiL. https://www.odil-projekt.de/de/startseite.html, Stand 28.06.2022.

Kliem, L.; Wagner, J.; Olk, C.; Keßler, L.; Lange, S.; Krachunova, T.; Bellingrath-Kimura, S.: Digitalisierung der Landwirtschaft. Schriftenreihe des IÖW, 222(22), 2022.

Koglin, I.; Dragon Dreaming. Projekt-Management für kollektive Kreativität und Nachhaltigen Erfolg. https://wwww.fuereinebesserewelt.info

Kramer, C.; Pfaffenbach, C.: Methoden der Raumanalyse, sozialwissenschaftliche. In Handwörterbuch der Stadt- und Raumentwicklung, Akademie für Raumordnung und Landesplanung, Hannover, 1487-1498, 2018.

Krüger, P. L: Datensouveränität und Digitalisierung: Probleme und rechtliche Lösungsansätze. Zeitschrift für Rechtspolitik, 49(7), 190-192, 2016.

Kuhn, L.; Bobojonov, I. B.; Glauben, T.: Landwirtschaft in Zeiten der Dürre: Wie Digitalisierung ein nachhaltiges Risikomanagement unterstützen kann. IAMO Policy Brief. No. 35, 4, 2018.

Landesamt für Statistik Niedersachsen, https://www.statistik.niedersachsen.de/presse/viehbestande-in-niedersachsen-endgultige-ergebnisse-der-landwirtschaftszahlung-2020-202032.html, Stand 13.06.2022.

Landschof, T.: Zu Digitalisierung, Datenschutz und Datensicherheit von Unternehmen. In Herausforderungen für Familienunternehmen, Vol. 7, Nomos Verlagsgesellschaft, Baden-Baden, 29-42, 2020.

Latacz-Lohmann, U.; Schreiner, J.: „Was wünschen Sich Landwirte von der Agrarpolitik Nach 2020?“. Agrar- und Ernährungswirtschaft zwischen Ressourceneffizienz und gesellschaftlichen Erwartungen, 55-66, 2018.

Lutz, K. J.: Digitalisierung der Landwirtschaft: Revolution mit evolutionärem Charakter. In CSR und Digitalisierung, Springer Gabler, Berlin, Heidelberg, 429-442, 2017.

Martin, R.: The design of business: why design thinking is the next competitive advantage. Harvard Business School Publishing, Brighton, 2009.

Martínez, J.: Chancen und Risiken der Digitalisierung in der Landwirtschaft – die rechtliche Dimension. Przegląd Prawa Rolnego, 2(19), 13-44, 2016.

Mayring, P.; Fenzl, T.: Qualitative Inhaltsanalyse. In: Methoden der empirischen Sozialforschung. 2., vollständig überarbeitete und erweiterte Auflage, Springer Fachmedien, Wiesbaden, 633-648, 2019.

Mohr, S.; Höhler, J.: Öffentliche Meinung zur Digitalisierung in der Landwirtschaft und ihren Auswirkungen. 40. GIL-Jahrestagung, Digitalisierung für Mensch, Umwelt und Tier, Gesellschaft für Informatik e.V., Bonn, 2020.

Nüssel, M.: Landwirtschaft 4.0 – die Waffe gegen Hunger und Umweltzerstörung? In Digitalisierung im Spannungsfeld von Politik, Wirtschaft, Wissenschaft und Recht, Springer Gabler, Berlin, Heidelberg, 343-363, 2018.

Ramesohl, S.; Gunnemann, A.; Berg, H. Digitalisierung gestalten - Transformation zur Nachhaltigkeit ermöglichen: eine Studie im Auftrag von Huawei Technologies Deutschland GmbH, Wuppertal Institut für Klima, Umwelt, Energie gGmbH, Wuppertal, 2021.

Reuter, C.; Schneider, W.; Eberz, D.: Resilient Smart Farming (RSF) – Nutzung digitaler Technologien in krisensicherer Infrastruktur. 39. GIL-Jahrestagung, Digitalisierung für landwirtschaftliche Betriebe in kleinstrukturierten Regionen - ein Widerspruch in sich? Gesellschaft für Informatik e.V., Bonn, 177-182, 2019.

Schawel, C.; Billing, F.: Top 100 Management Tools. Das wichtigste Buch eines Managers - Von ABC-Analyse bis Zielvereinbarung. 5te, überarbeitete Auflage, Springer Fachmedien, Wiesbaden, 321, 2014.

Scheuren, S.; Bauer, J.; Gehrs, F.; Grothaus, H.-P.: Offene Software-Plattform für Dienstleistungsinnovationen in einem Wertschöpfungsnetz in der Landwirtschaft. In Dienstleistungsinnovationen durch Digitalisierung. Springer Gabler, Berlin, Heidelberg, 483-531, 2021.

Schön, H.: Elektronik und Computer in der Landwirtschaft. Ulmer Eugen Verlag, Stuttgart, 1993.

Schmidt, C.: Landwirtschaft 4.0 – Digitalisierung als Chance für eine nachhaltige Landwirtschaft. In Digitalisierung im Spannungsfeld von Politik, Wirtschaft, Wissenschaft und Recht, Springer Gabler, Berlin, Heidelberg, 397-407, 2018.

Serrat, O.: Design Thinking. In Knowledge Solutions. Springer, Singapore, 2017.

Stiene, S.; Scheuren, S.; Günther, M.; Lingemann, K.; Lauer, A.; Bernardi, A.; Hertzberg, J.: Architektur einer offenen Software-Plattform für landwirtschaftliche Dienstleistungen. Informatik in der Land-, Forst-und Ernährungswirtschaft, Gesellschaft für Informatik e.V., Bonn, 141-144, 2017.

Stein, P.: Forschungsdesigns für die quantitative Sozialforschung. In Handbuch Methoden der empirischen Sozialforschung. Springer VS, Wiesbaden, 135-151, 2014.

Umstätter, C.; Martini, D.; Adrion, F. Opinion Paper: Digitales Tiermonitoring− Was bringt die Zukunft. Landtechnik, 75(1), 14-23, 2020.

Van Berkel, K.; Manickam, A., Wicked World - Complex Challenges and Systems Innovation, Taylor & Francis, London, 2020.

VDI Statusreport: Industrie-4.0-Technologien in der Landwirtschaft, VDI Darmstadt, 2021.

Von Unger, H.: Partizipative Forschung. Einführung in die Forschungspraxis. Springer Fachmedien, Wiesbaden, 2014.

Witzel, A.: Das problemzentrierte Interview. Forum Qualitative Sozialforschung, 1(9), 2002.

Digitale Wandlungsprozesse: 30 Jahre Bergbausanierung im Spiegel der Umweltdatenerhebung

Jana Götze[1], Wilko Hinz[2], Mario Päßler[3], Axel Hiller [4]

Abstract: Die Wismut GmbH ist ein Unternehmen zur Sanierung der Hinterlassenschaften des Uranerzbergbaues. Seit nunmehr 30 Jahren werden so genannte Sanierungsobjekte (Halden, industrielle Absetzanlagen, Grubengebäude u. a.) in Sachsen und Thüringen saniert und einer Nachnutzung zurückgeführt. Derzeit erfolgt der Übergang von der Kernsanierung zu den Langzeitaufgaben, die bis 2050 geplant sind. Ohne bestehende digitale Strukturen, die es auch in Zukunft weiterzuentwickeln gilt, ist dieser Wandel nicht zu meistern. Auf dem Gebiet des Wissensmanagements bedient sich die Wismut GmbH einer Auskunftsplattform namens AL.VIS/W, welches Daten – und Informationen zu Naturwissenschaft & Technik vereint. Der Beitrag widmet sich den digitalen Wandlungsprozessen, die im Zuge der Bergbausanierung stattgefunden haben. Am Beispiel der Erhebung radiologischer Umweltdaten soll einer der tragenden Datenworkflows aufgezeichnet werden, der exemplarisch für viele weitere Routinen steht, die im Unternehmen gepflegt und gelebt werden.

Keywords: Uranbergbausanierung, Dateninformationssystem, Renaturierungsprojekt

1 30 Jahre Bergbausanierung und ihre digitale Abbildung

Die Wismut GmbH ist ein Unternehmen zur Sanierung der Hinterlassenschaften des Uranerzbergbaues. Seit nunmehr 30 Jahren werden Halden, industrielle Absetzanlagen, Tagebaue und Grubengebäude sowie zugehörige Betriebsflächen und Anlagen in Sachsen und Thüringen saniert bzw. einer Nachnutzung zurückgeführt. Im nationalen wie auch internationalen Vergleich ist diese Aufgabe als Großprojekt im Bereich Bergbausanierung und Umweltschutz einzuordnen. Die physischen Sanierungsarbeiten werden voraussichtlich bis zum Jahr 2028 abgeschlossen sein. Derzeit erfolgt der Übergang von der Kernsanierung zu den Langzeitaufgaben, die bis 2050 geplant sind. Dieser Wandel wird neben den verbleibenden Sanierungstätigkeiten begleitet von Optimierungen bestehender Workflows, der Kompensation von Wissensverlusten aufgrund des zunehmenden Ausscheidens von Personal sowie dem Transfer von Wissen an die neu heranwachsende Generation. Ohne bestehende digitale Strukturen, die es auch in Zukunft weiterzuentwickeln gilt, ist dieser Wandel nicht zu meistern. Auf dem Gebiet des Wissensmanagements bedient sich die Wismut GmbH einer Auskunftsplattform namens AL.VIS/W, welches Daten und In-

[1]Wismut GmbH, Abteilung Informationsmanagement, Jagdschänkenstr. 29, 09117 Chemnitz, j.goetze@wismut.de
[2] Wismut GmbH, Abteilung Wassermanagement/Monitoring, Jagdschänkenstr. 29, 09117 Chemnitz, w.hinz@wismut.de
[3] Wismut GmbH, Projekt Umweltmessung, Talstr. 7, 08118 Hartenstein, m.paeszler@wismut.de
[4] Wismut GmbH, Abteilung Informationsmanagement, Talstr. 7, 08118 Hartenstein, a.hiller@wismut.de

doi:10.18420/inf2022_69

formationen aus dem Bereich Naturwissenschaft und Technik vereint. Es bietet unternehmensübergreifend einen einheitlichen browserbasierten Webzugriff auf Datenbanken der folgenden naturwissenschaftlichen Bereiche: Hydrologie, Klimatologie, Geotechnik, Strahlenschutz, und Umweltanalytik [Go17]. Neben dem Komplex der Umweltmessdaten stehen weitere Informationsquellen zur Verfügung, die beispielsweise in Form von standardisierten Dokumentationen Objekte, Aktivitäten und Ergebnisse der Bergbausanierung beschreiben. Ein integriertes WEB-GIS System ermöglicht die übergreifende Recherche und Visualisierung aus dem Komplex der Umweltmessdaten sowie den zur Verfügung stehenden GIS-Ebenen (von der einfachen Punktkoordinate über das Orthophoto bis zum Risswerk). Der Wissensverbund wird zwischen den einzelnen Informationskomplexen via Suchanfrage über Geometrien, ID-Nummern, Semantik usw. hergestellt. Der Beitrag widmet sich den digitalen Wandlungsprozessen die im Zuge der Bergbausanierung stattgefunden haben. Am Beispiel der Erhebung radiologischer Umweltdaten soll einer der tragenden Datenworkflows aufgezeichnet werden, der exemplarisch für viele weitere Routinen steht, die im Unternehmen gepflegt und gelebt werden. Dabei darf der Begriff „Informatik“ bzw. „Digitalisierung“ nicht erst an der Datenbank und ihrer Visualisierung beginnen, sondern steht am Anfang eines Konzeptes, welches den Weg der Daten bis heute beschreibt.

2 Umweltdatenerhebung im digitalen Wandel

Um die bestehenden digitalen Strukturen, deren Workflows sowie den derzeitigen Gesamtdatenbestand zu verstehen und evaluieren zu können, bedarf es eines Blickes sowohl in die Vergangenheit als auch in die abteilungsübergreifende Arbeit. Dieser Gesamtblick soll anhand der Radonkonzentration erfolgen (nachstehend als Parameter Radon bezeichnet). Radon dient als Grundlage zur Dosisabschätzung für die Bevölkerung sowie als Entscheidungsgrundlage über einen notwendigen Sanierungsbedarf. Daten dazu werden seit Gründung der Wismut GmbH im Jahr 1990 gemessen. Die ersten Messungen erfolgten anhand einer flächendeckenden „Bestandsaufnahme, zur Klärung der stofflichen Auswirkung des Uranerzbergbaues und der Uranaufbereitung auf die Umwelt unter besonderer Berücksichtigung der radiologischen Situation“ [Ha92]. Auf Grundlage des nunmehr bekannten radiologischen Inventars gaben die Strahlenschutzverordnung sowie die Richtlinie zur Emissions- und Immissionsüberwachung den rechtlichen Rahmen vor, an welchen Objekten, Flächen und Umweltkompartimenten (Luft, Wasser, Boden, Organismen) welches Parameterspektrum, neben Radon zu überwachen, zu dokumentieren und zu berichten ist. Aufbauend auf dem rechtlichen Rahmen wurden interne Konzepte abgeleitet, die konkrete Handlungsanweisungen vorgaben, wie häufig, in welchem Maß, und in welcher Art und Weise die Datenablage und -weiterverarbeitung erfolgen sollte. Die Datenhaltung war bis 1995 über die damaligen Standorte dezentral und dateibasiert organisiert. Neben dem rechtlichen Rahmen entwickelte sich die Interpretation der Umweltüberwachung von der einfachen Bestandsaufnahme zu komplexen Messprogrammen, die das stetig anwach-

sende Datenvolumen mitbestimmten. Grundsätzlich wird der Parameter Radon für die Bewertung des Einflusses auf die Bevölkerung, der Sanierungsobjekte und des Arbeitsortes herangezogen. Dabei wird unterschieden in Basis-, Sanierungs- und Arbeitsplatzmonitoring. Der stete Zuwachs an Daten sowie deren Relevanz für diverse Messprogramme bedingte ab 1995 die Einführung einer zentralen Umweltdatenbank auf Oracle-Basis. Der Import sowie die Qualitätssicherung der Daten, hinsichtlich von Ausreißern bzw. festgelegten Plausibilitätsprüfungen, wurden durch clientbasierte dezentrale und messprogrammbezogene Einzelsoftware organisiert. Das Ausmaß der Beeinflussung durch den Parameter Radon (Bevölkerung, Sanierungsobjekte und Arbeitsort) bestimmte die einzusetzende Messtechnik. Dabei unterliegen die **Technischen Entwicklungen** neben der eigentlichen Messaufgabe ebenso verändernden Bedingungen (Kosten, Verfügbarkeit, Support), die maßgeblich das spätere Erscheinungsbild der nunmehr vorliegenden digitalen Strukturen geprägt haben. Die Radonüberwachung erfolgt messprogrammbezogen (Tab. 1: Charakteristik verschiedener Messprogramme).

Messzyklus	**Messfrequenz**	**Primärdaten**	**Messeinrichtungen**
saisonal	Winter- und Sommermess-kampagne	integrative Daten	stationär
kontinuierlich	Zeitaufgelöste 10-Minuten bis Stundenwerte	Einzelwerte	transportabel
stichprobenartig	Definierter Kurzzeitraum	Integrative Einzelwerte	transportabel

Tab. 1: Charakteristik verschiedener Messprogramme

Das Basismonitoring untersucht standortbezogen den Einfluss auf die Bevölkerung und Umgebung und verwendet hierfür Messwerte in geringer zeitlicher Auflösung, dafür in breiterer Verteilung an stationären Randflächen der Sanierungsobjekte. Das Sanierungsmonitoring benötigt im jeweiligen Betrachtungszeitraum eine höherfrequente Auflösung der Messwerte, um etwaige Effekte oder Erfolge aufzuzeigen, die mit dem räumlich fortschreitenden Sanierungsarbeiten flexibel, transportabel einzurichten sind. Für den Arbeitsort werden Messwerte mit der höchsten Frequenz jedoch integrativ über den Arbeitstag herangezogen. Messwerte zum Parameter Radon benötigen allein schon hierfür verschiedene Ablage-/Speichermöglichkeiten und Visualisierungsformen: Sei es die Darstellung des Messwertes zum Zeitpunkt „x“ oder im Zeitraum „y“, die Zugehörigkeit zu unterschiedlichen Messprogrammen bzw. die visuell getrennte Vergleichbarkeit hinsichtlich rechtlicher Anforderungen. In 2005 endete die clientseitige Einzelimportlösung. Nunmehr war eine standort- und messprogrammübergreifende Softwarelösung geschaffen worden. Die Ausprägung und Angleichung digitaler Strukturen über die Zeit hinweg unterliegt für

das Beispiel Radon einem rechtlichen Rahmen, den daraus resultierenden Umweltüberwachungsprogrammen sowie der erforderlichen Messtechnik (Abb. 1: prägende Rahmenbedingungen digitaler Strukturen). Sie bedingen sich untereinander und wurden parallel miteinander abteilungsübergreifend entwickelt. Dazu bedarf es einer abgestimmten Zusammenarbeit zwischen den Datenproduzenten über die Datenverwaltung bis hin zum Datenkonsumenten.

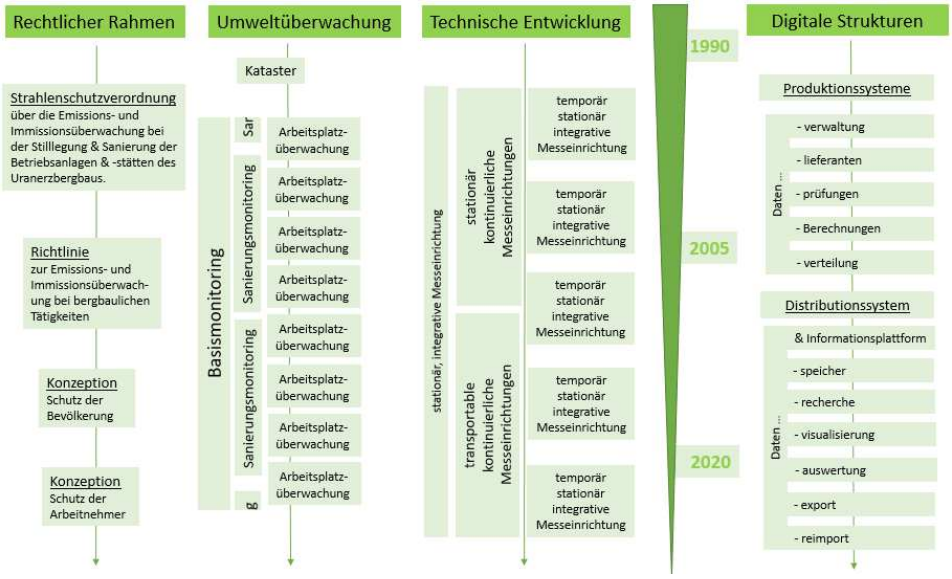


Abb. 1: prägende Rahmenbedingungen digitaler Strukturen

Nicht nur die oben beschriebenen Rahmenbedingungen prägen die digitalen Strukturen - die Weiterentwicklung von Softwarekomponenten und -produkten selbst wirkt ebenso bestimmend auf das Gesamtsystem ein. Mit Einstellung von Entwicklung und Support der bestehenden Einzelimportlösungen aus 2005 erfolgte zwingend ein Umdenken ab 2015. Dies galt nicht nur für den beschriebenen radiologischen Parameter Radon. Die Entscheidung fiel auf ein neues Softwarepaket, welches per Lizenz (TimeSeries; Wisutec Umwelttechnik Zweigniederlassung der G.E.O.S. Ingenieurgesellschaft mbH) erworben wurde. Ziel war es, verschiedene Messeinrichtungen (automatisiert, manuell, hybrid) auf einer zentralen Plattform anzubinden und direkt zu verwalten. Im Ablösungsprozess wurden Routinen hinsichtlich ihrer zukünftigen Verwendung und Funktion geprüft. Somit konnten bisherige Arbeitsroutinen verschlankt oder für den derzeitigen Verwendungszweck optimiert werden. Die Herausforderung hierbei war, die vielfältige Themenspezifik jeder einzelnen Vorgängeranwendung in einem Gesamtsystem wiederzufinden, ohne dass die Themen sich in den angebotenen Funktionen gegenseitig behindern. Somit sind nunmehr standortübergreifend Messstationen der naturwissenschaftlichen Bereiche Radiologie, Hydrologie, Klimatologie angebunden.

3 Das Konzept, welches den Weg der Daten beschreibt

Die Strategie des Datenmanagements bei Wismut besteht in der Gliederung der digitalen Strukturen in ein Produktions- und ein Distributionssystem. Der Vorteil dieses Ansatzes besteht in der Möglichkeit, sich flexibel auf den sich dynamisch verändernden Datenbestand einstellen, kostenintensive Einzellizenzen einsparen und ein kostengünstiges Arbeitsmittel zur firmenübergreifenden Nutzung bereit stellen zu können. Die Produktionssysteme verwalten alle Zu- und Abgänge von Messpunkten, Datenlieferanten, Qualitätssicherungsprüfungen sowie die Weiterverarbeitung zu möglichen Sekundärdaten (Statistik/Interpolation) und deren Verteilung in das Distributionssystem. Der Zugang zu dieser Ebene ist abteilungsspezifisch spezialisierten Nutzerkreisen zugänglich und enthält Daten vom Messwert bis hin zum technischen Hilfsparameter. Die für den Endnutzer relevanten Daten werden periodisch an das Daten- und Informationssystem AL.VIS/W übergeben, zentral verwaltet und über eine Weboberfläche allen Mitarbeitern des Unternehmens zur Verfügung gestellt (Abb. 2: Daten- und Informationssystem AL.VIS/W – Ein Blick auf die Startseite).



Abb. 2: Daten- und Informationssystem AL.VIS/W – Ein Blick auf die Startseite

Dieses System besteht aus einer Oracel-Datenbank, welche in ihrem Datenmodell alle wesentlichen Daten, unternehmensweit speichert. Die Visualisierungs-, Recherche-, Auswertungs- und Exportwerkzeuge sind innerhalb von über 20 Anwendungen (Wisutec Umwelt-

technik GmbH mit Projektleitung der Wismut) themenspezifisch verteilt und liegen eingebettet in der Lizenzkomponente Cardo, welche den kartographischen Zugang, die Applikationssteuerung sowie die Nutzerverwaltung enthält (Firma IDU Zittau). Der Speicherplatz der erhobenen Messwerte des Parameters Radon ist die Anwendung „Luft“. Die Anwendung „Luft“ beinhaltet somit ein definiertes Datenmassiv, dessen strukturierte Ablage bestimmt wurde durch die gegebenen Rahmenbedingungen. Die Datenbank stellt nicht nur ein Arbeitsmittel dar, sondern reflektiert in ihrer Struktur auch die stete, bedarfsgerechte Anpassung an das Datendargebot (Abb. 3: Weg der Daten am Beispiel des Parameters Radon).

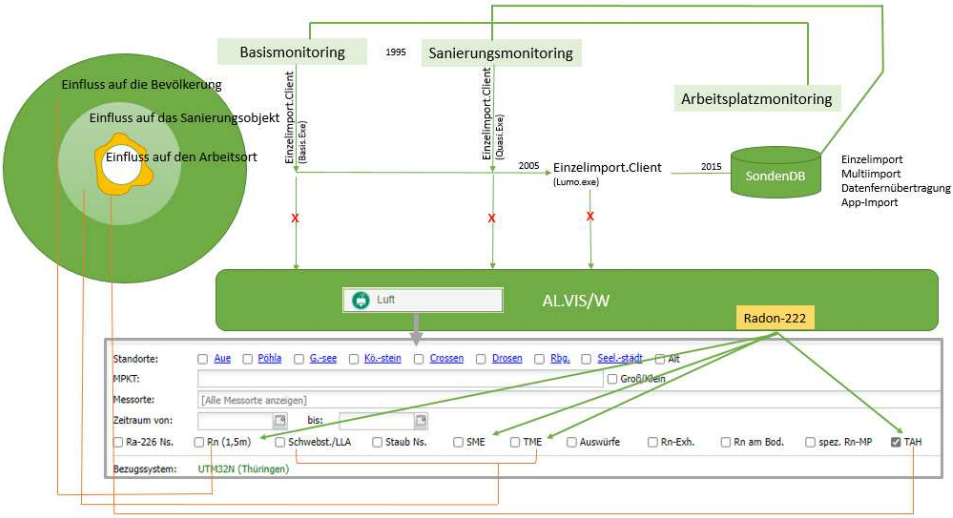


Abb. 3: Weg der Daten am Beispiel des Parameters Radon

In alljährlichen anwenderspezifischen Schulungen wird dieses Wissen um die Historie und den Datenverbund an die Nutzer weitervermittelt.

4 Zusammenfassung und Ausblick

Die aufgezeichnete Historie zum Parameter Radon zeigt, dass innerhalb von 30 Jahren Bergbausanierung die Digitalisierung als notwendiges Werkzeug unabdingbar war. Sie dient nicht nur als tägliches Arbeitsmittel, sondern sie begleitet den fortwährenden Planungs- und Entscheidungsprozess während der Sanierungstätigkeit. Ein nächstes Entwicklungsziel besteht darin, nach Abschluss der Kernsanierung, das Daten- und Informationssystem AL.VIS/W fit zu machen für die kommenden Aufgaben (Langzeitmonitoring, Integration von analogen Beständen, …) und für Nutzer ohne Vorkenntnisse aus dem praktischen Sanierungsablauf. Die Bedarfsträger für künftige Fragestellungen werden neben Mitarbeitern der Wismut GmbH auch potentielle Nutzer aus Planungsbüros, Eigentümer

sanierter Flächen u. a. sein. Wie eingangs erwähnt, fängt die Digitalisierung nicht erst an der Softwarelösung selbst an, sondern bereits im konzeptionellen Stadium. Neben den genannten Rahmenbedingungen und dem sich daraus entwickelnden Konzept bedarf es finanzieller und personeller Untersetzung, um den Digitalisierungsweg kontinuierlich und schrittweise konzipieren, begleiten, aufarbeiten und weitergeben zu können.

5 Literaturverzeichnis

[Go17] Götze, J.; Kreyßig E., et al.: Management von Daten im Rahmen des Wismut-Umweltprojektes. Lecture Notes in Informations, Bonn, 2017

[Ha92] Hähne, D.; et al.: Ergebnisbericht Wismut-Umweltkataster. interner Bericht, 1992

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 831

Prototypische Implementierung eines Digital Twin und Predictive Maintenance Methoden am Beispiel des Wasserwerks Langeoog

Marius Wybrands[1] Sven von Höveling[2] Marco Muras[3] Heinrich Töpfer[4] Jorge Marx Gómez[5]

Abstract: In der Wasserwirtschaft werden vermehrt die Potenziale von Digital Twin und Predictive Maintenance diskutiert. Es fehlt jedoch an Prototypen, welche die theoretischen Überlegungen mit den Bedarfen der Praxis abgleichen. Um diese Lücke zu schließen, wurden am Beispiel des Wasserwerks der Nordseeinsel Langeoog die Konzepte Digital Twin und Predictive Maintenance erprobt. Das Ziel war es, einen horizontalen Prototypen zu entwickeln, der vom Sensor bis zum Dashboard alle Informationsebenen abbildet, um Rückschlüsse auf Potenziale zur Digitalisierung und Einschränkungen in den Technologien zu ermitteln. Durch den Prototypen konnte gezeigt werden, dass durch den Einsatz von Digital Twins und Predictive Maintenance Methoden ein besseres Verständnis über den aktuellen Systemzustands für das Wartungspersonal erlangt werden kann.

Keywords: Digital Twin; Wasserversorgung 4.0; Wasserwerk; Predictive Maintenance

1 Motivation

Durch die Digitalisierung ausgelöst, werden Unternehmensprozesse neu strukturiert und an die fortführende Digitalisierung angepasst. Dies führt zu tiefgreifenden Veränderungen, die Unternehmen vor neue Herausforderungen stellen. So auch die Unternehmen im wasserwirtschaftlichem Sektor [Ho19, S. 5]. Die Zusammenführung von Daten aus unterschiedlichen Quellen wird, im Zuge der veröffentlichten Handlungsempfehlung des deutschen Umweltbundesamtes, als ein zentrales Handlungsfeld für die Entwicklung der Wasserwirtschaft 4.0 hervorgehoben. Dies gilt ebenso für die weitere Verarbeitung von Daten durch Simulationen, Modellierungen und Live-Visualisierungen auf einer geeigneten Plattform [Ho19, S. 97 ff.].

[1] Carl von Ossietzky Universität Oldenburg, Wirtschaftsinformatik/VLBA, Ammerländer Heerstraße 114–118, 26129 Oldenburg, Deutschland, marius.wybrands@uol.de
[2] Carl von Ossietzky Universität Oldenburg, Wirtschaftsinformatik/VLBA, Ammerländer Heerstraße 114–118, 26129 Oldenburg, Deutschland, sven.von.hoeveling@uol.de
[3] Carl von Ossietzky Universität Oldenburg, Wirtschaftsinformatik/VLBA, Ammerländer Heerstraße 114–118, 26129 Oldenburg, Deutschland, marco.muras@uol.de
[4] Carl von Ossietzky Universität Oldenburg, Wirtschaftsinformatik/VLBA, Ammerländer Heerstraße 114–118, 26129 Oldenburg, Deutschland, heinrich.toepfer@uol.de
[5] Carl von Ossietzky Universität Oldenburg, Wirtschaftsinformatik/VLBA, Ammerländer Heerstraße 114–118, 26129 Oldenburg, Deutschland, jorge.marx.gomez@uol.de

doi:10.18420/inf2022_70

Die Errungenschaften der angehenden Digitalisierung und der stetigen Weiterentwicklung von Modellen aus dem Bereich Machine Learning ermöglichen es, Wartungen von Pumpen oder weiteren technischen Komponenten prädiktiv vorzunehmen [Sc20]. Bei Predictive Maintenance (PdM) sollen dabei Fehler proaktiv verhindert werden, indem Fehleranzeichen früh erkannt und gezielte Gegenmaßnahmen eingeleitet werden, um so unerwartete Ausfälle zu vermeiden, Zuverlässigkeit zu erhöhen und Betriebskosten zu reduzieren [Ra19, S. 1].

Ein Digital Twin (DT) ist dabei das virtuelle Abbild eines physischen Objektes und ermöglicht den automatischen Datenfluss zwischen dem physischen und digitalen Objekt. Somit können Daten und Steuerungsinformationen übermittelt werden, wobei sich die beiden Objekte gegenseitig beeinflussen. So können z. B. frühzeitig (problematische) Systemzustände vorhergesagt werden und direkt gegengesteuert werden [Gü20, S. 5][Kr18, S. 1017][Fu20, S. 108953].

Aus dem Bedarf an neuen Erkenntnissen im Bereich Verarbeitung von Daten durch Simulationen, Modellierungen und Live-Visualisierung wurde betrachtet, welche möglichen Einsatzpotenziale es für DT und PdM in der Wasserwirtschaft gibt. Mit dem Praxispartner OOWV (Oldenburgisch-Ostfriesischer Wasserverband) wurde das Wasserwerk auf der Nordseeinsel Langeoog als Anwendungsfall ausgewählt. Dieser konkrete Anwendungsfall bringt einige Besonderheiten mit sich. Zum einen sind Wartungsarbeiten durch Personal vom Festland aufwendiger, was den Stellenwert von Predictive Maintenance nochmal erhöht. Zum anderen ist das Wasserwerk das einzige in einem mehr oder weniger abgeschlossenem System, weshalb das Thema Ausfallsicherheit umso wichtiger ist. Außerdem gibt es aufgrund des Tourismus starke Unterschiede in der Auslastung im Jahresverlauf, sodass im Winter auf kleinere Pumpen umgestellt wird. Der Anwendungsfall wurde nicht explizit aufgrund dieser Besonderheiten ausgewählt, eignet sich jedoch gut, da es, auch aufgrund der Besonderheiten, große Einsatzpotenziale für DT und PdM gibt.

Folgende Fragestellungen wurden für den Anwendungsfall festgelegt:

1. Inwiefern kann das Wasserwerk von Langeoog mithilfe von DT abgebildet werden?
2. Welche Instandhaltungsstrategien und PdM-Methoden eignen sich, um eine Instandhaltung für das Wasserwerk zu planen?
3. Wie können DT und PdM-Methoden in diesem Kontext kombiniert werden?

In einem webbasiertes Expertensystem wurden prototypisch die Konzepte eines DT und PdM kombiniert und mithilfe eines Frontends alle wesentlichen Informationen für den Experten angemessen darstellt. Ein Expertensystem ist ein System, welches Fachwissen aus einer bestimmten Wissensdomäne bereitstellt und verwaltbar macht. Da zur Bedienung des Systems eine gewisse Expertise vorausgesetzt wird, sind die Benutzer in der Regel Fachexperten, die vom System die nötigen Informationen zur Erklärung von Lösungsvorschlägen erhalten. Weiterhin soll die Wasserinfrastruktur von Langeoog mithilfe von DT abgebildet

werden. Hierzu soll das zu entwickelnde System in den Instandhaltungsprozess im wasserwirtschaftlichen Kontext des lokalen Betreibers integrierbar sein. Als Betrachtungseinheit wird sich auf das wasserwirtschaftliche System des Wasserwerks auf Langeoog beschränkt.

2 Prototypische Implementierung

Die Architektur wurde mittels Microservices umgesetzt und in das bereits bestehende wasserwirtschaftliche Informationssystem WISdoM [Wy21] integriert, welches ebenfalls auf eine Microservice-Architektur setzt. Das Expertensystem besteht aus fünf Kernkomponenten, die unterschiedliche Funktionen erfüllen. Alle Services greifen auf den Data-Service zu, um über die REST-Schnittstelle des Data-Services auf die Datenbank zugreifen zu können. Die Schnittstellen stellen verschiedene HTTP-Methoden bereit, damit über diese Methoden CRUD-Operationen auf der Datenbank für unterschiedliche Daten ausgeführt werden können. Zum Speichern von verschiedenen Daten wie Sensorwerten oder Stammdaten von Anlagen, wie bspw. Pumpen, wird eine PostgreSQL-Datenbank verwendet.

In Abbildung 1 sind die Datenströme innerhalb des Expertensystems dargestellt. Zudem wird der Kommunikationsfluss der Sensordaten bis zum Frontend aufgezeigt. Die Sensoren, die an den Pumpen auf Langeoog angebracht wurden, sind sogenannte FAG SmartChecks. Mittels der proprietären Lösung Talk2M des Unternehmens EWON (HMS Networks) wurde eine VPN-Verbindung zu den Sensoren im Wasserwerk aufgebaut. Sobald der Connector-Service die Daten der Sensoren abgefragt hat, werden sie mittels des Messagebrokers Apache Kafka als Datenstrom weiterverarbeitet. Auch alle anderen Services können auf den dargestellten Datenstrom zugreifen und Daten lesen oder schreiben.

Kafka erlaubt den bidirektionalen Austausch sowie das Speichern und Verwalten von Daten bzw. Datenströmen. Zudem ist es möglich, Datenströme an mehrere Systeme oder Services weiterzugeben. Dadurch wird mittels Kafka ein zentraler Dienst für alle Datenströme bezüglich Sensorwerte, Vorhersagen, Anomalien oder Events bereitgestellt. Sollten neue Daten, beispielsweise durch Sensoren, verfügbar sein, so werden diese mittels Kafka verteilt. Alle ein- und ausgehenden Datenströme von Kafka werden in Abbildung 1 rot dargestellt.

Um die bidirektionale Verbindung zwischen Frontend und dem Data-Service herstellen zu können, wird das WebSocket-Protokoll verwendet. Hiermit wird das Frontend über neue Daten zeitnah informiert. Zusätzlich wird für das Frontend eine Möglichkeit benötigt, die vorhandenen Daten aus den Kafka-Topics anzufordern, um den Datenstrom zu vervollständigen. Die Möglichkeit, eine bidirektionale Verbindung zu etablieren, ist die wesentliche technische Erweiterung zu dem in Wybrands et al. [Wy21] vorgestellten wasserwirtschaftlichen Informationssystem.

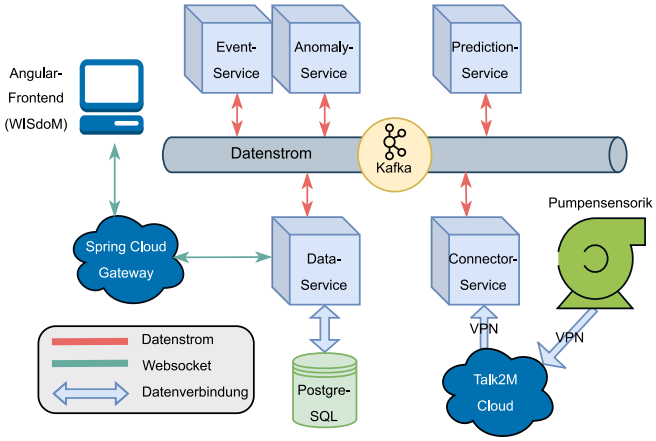


Abb. 1: Dargestellt sind die INDICATE-Architektur und die Datenströme innerhalb dieser Architektur. Eine Besonderheit sind die Pumpensensorik und das Abgreifen deren Daten über die Talk2M Cloud. (Quelle: Eigene Darstellung)

2.1 Predictive Maintenance Service

Das Konzept PdM wird im Expertensystem im PDM-Service mithilfe eines Prediction-Services und eines Anomalie-Services realisiert. Der Prediction-Service dient der Prognose von Messeinheiten der Sensorik mithilfe von Modellen. Der Anomalie-Service setzt Anomalie-Modelle zur Erkennung von Ausreißern in Live- und historischen Daten ein. Die PdM-Services nutzen eine REST-Schnittstelle, um auf die historischen Daten der Reinwasserpumpe vom Data-Service zuzugreifen, sowie Confluent Kafka, um auf den realisierten Datenstrom zugreifen zu können. Der Anomalie-Service arbeitet im Aufbau nahezu analog zum Prediction-Service und neben der Auswahl des verwendeten Anomalie-Modells und den Update-Frequenzen ist auch die Wahl der Untergrenze für die Klassifizierung der Daten anhand des Anomalie-Scores konfigurierbar. Abbildung 2 zeigt die Anzeige von Ergebnissen des Prediction- und Anomalie-Services im Diagramm des Expertensystems.

2.1.1 Data-Service

Der Data-Service ist eine Kernkomponente innerhalb des Expertensystems und wurde mithilfe des Frameworks Spring realisiert. Er hat die Aufgabe allen Services sowie dem Frontend über eine REST-API Zugriff auf die Datenbank zu gewähren. Zudem stellt der Data-Service eine WebSocket-Schnittstelle für das Frontend bereit. Alle Datenmodelle werden mittels Hibernate und PagingAndSortingRepositorys den anderen Diensten zur Verfügung gestellt. Neben den REST-Schnittstellen ist der Data-Service auch an Kafka

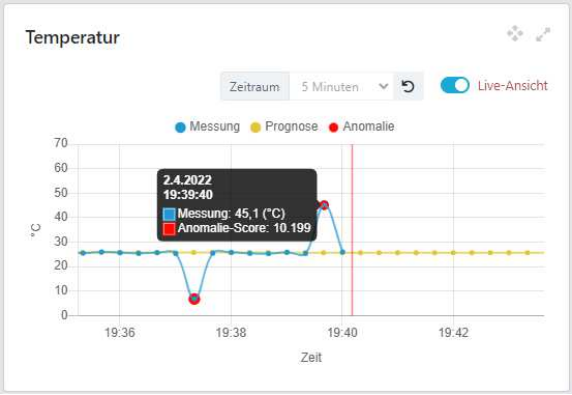


Abb. 2: Dargestellt sind die Ergebnisse des Prediction-Services in Form der Zeitreihenvorhersage (gelb) sowie des Anomalie-Services in Form des Anomalie-Scores (rot). Für die Erzeugung der Darstellung wurde ein Datengenerator verwendet, weshalb die konkreten Werte keine Aussagekraft haben.

angebunden. Der Service bekommt darüber sowohl die Werte der verschiedenen Sensoren über die konfigurierten Ports, als auch Events, die von unterschiedlichen Services gemeldet werden können. Zusätzlich ist der Service an die PostgreSQL-Datenbank angebunden. Der Data-Service ist der einzige Service, der die Datenbank bearbeiten kann. Alle anderen Services müssen für den Zugriff auf die Datenbank den Data-Service kontaktieren und die bereitgestellten Schnittstellen nutzen.

2.1.2 Connector-Service (Smartcheck-Connector)

Der Smartcheck-Connector ist ein entwickelter Service, um die Messdaten der FAG Smartcheck Sensorik abzufragen. Der Smartcheck misst zyklisch verschiedene Messgrößen. Die Messdaten können dabei über die Webschnittstelle vom Webserver des Smartchecks ausgelesen werden. Mittels der Webschnittstelle werden die Messdaten nur per Pollingmechanismus abgefragt. Der Smart-check-Connector fragt alle zwanzig Sekunden zyklisch die aktuellen Messwerte der Sensoren ab, transformiert sie und schreibt die Ergebnisdaten per Pushmechanismus auf Kafka Topics.

2.1.3 Event-Service

Der Event-Service wird genutzt, um alle Events zu melden, die nicht von anderen Services gemeldet werden. Dazu gehören zum Beispiel Grenzwertüberschreitungen oder der Ausfall eines anderen Microservices. Die Datenstruktur sowie die Schnittstellen von Events werden durch den Data-Service bereitgestellt. Der Data-Service bietet zusätzlich ein

Konfigurationssystem, über den der Event-Service konfiguriert werden kann. So kann z. B. ein Grenzwert eingestellt werden. Sobald ein Event auftritt, wird es über eine entsprechende Kafka-Topic gemeldet.

2.2 Frontend

Basierend auf die in der Anforderungserhebung herausgearbeiteten Funktionsbereiche, wurden mehrere eigenständige Komponenten mittels Services entwickelt, welche der Abbildung 4 entnommen werden können. In den Services sind die Methoden zur Kommunikation mit dem Data-Service via REST bzw. WebSocket implementiert. Zur Strukturierung des Projektes werden neue Ordner für die Komponenten, welche grün markiert sind, angelegt sowie thematisch zusammenhängende Komponenten in Oberordnern zusammengefasst.

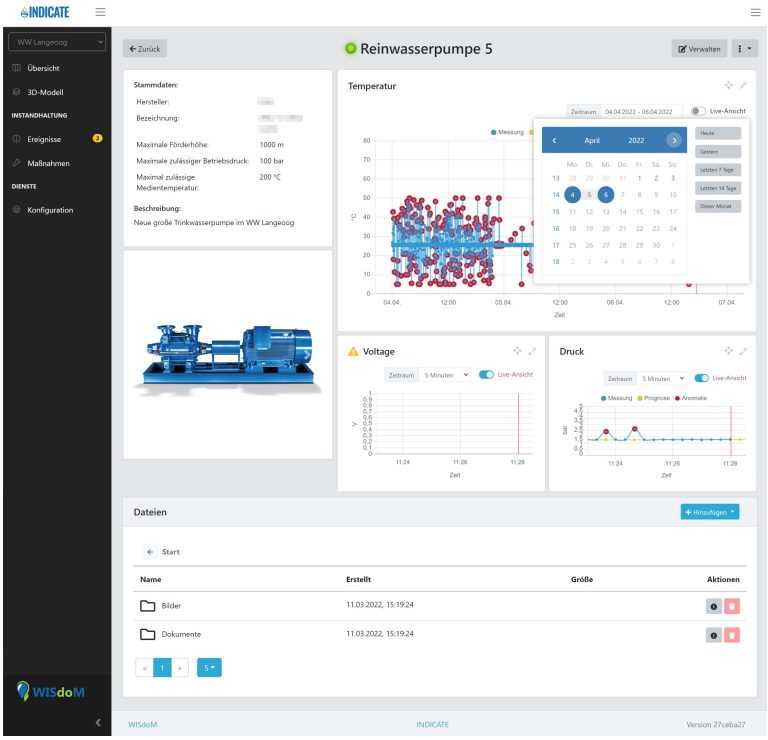


Abb. 3: Anlagenübersicht mit Bildergalerie, Dateisystem und mit Sensor- sowie Analysedaten

Für den Bereich Stationen (z. B. Wasserwerke oder Förderbrunnen) werden zwei Komponenten benötigt. Zum einen die Stationsübersicht, welche den Experten einen Überblick geben soll in Form einer Kartendarstellung mit Markierung und einer Auflistung der jeweiligen Standorte. Zum Anderen wird für den Administrator eine Komponente für die Verwaltung benötigt, um eine neue Station anlegen oder bestehende Stationen bearbeiten und löschen

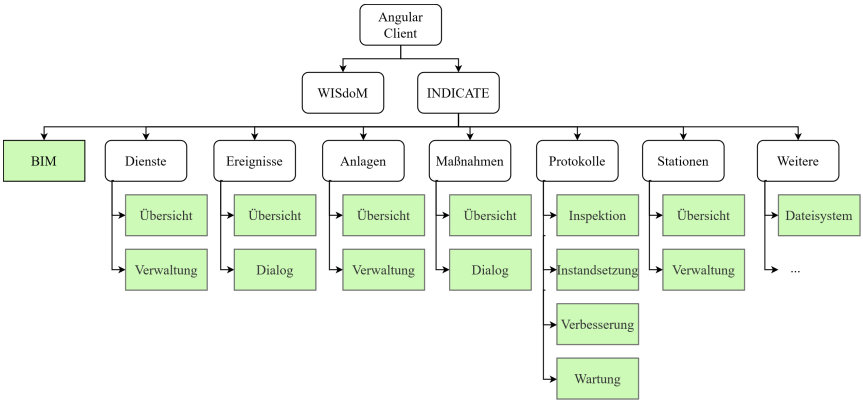


Abb. 4: Abstrahierte Übersicht der Komponenten im Expertensystem. Grün markierte Elemente spiegeln die Komponenten wider (Quelle: Eigene Dartstellung)

zu können. Auch für die Anlagen wird eine Anlagenübersicht benötigt zur Anzeige der Stammdaten, vorhandener Bilder, des Status der Anlage sowie hinterlegter Dateien. Damit der Experte zur Bewertung des Gesundheitszustandes der Anlage einen visuellen Überblick über die Messwerte und die prognostizierten Werte sowie Anomalien erhält, werden diese in Form von Diagrammen angezeigt (Abbildung 3), bei denen die Zeitintervalle per vordefinierten Zeiträumen oder intuitiv via Zoom und Verschieben angepasst werden können. Bzgl. der Steuerung wird es dem Experten ermöglicht, die Anlage ein- und auszuschalten. Da für die Realisierung eines DTs digitale Ein- und Ausgänge benötigt werden, um Daten vom digitalen Objekt zum physischen Objekt senden und auch von diesem empfangen zu können, wird in der Anlagenverwaltung neben der Verwaltung allgemeiner Daten und den Stammdaten u. a. auch das Verwalten dieser Schnittstellen integriert, welche mit zuvor genannten Diensten verknüpft werden, um Daten von diesen ins Frontend zu bringen bzw. vom Frontend für die Steuerung empfangen zu können.

Sofern ein Building Information Modeling (BIM) vorhanden ist, kann dieses sowohl in der Stationsübersicht als Orientierungshilfe, als auch in einer eigenen Seite mit weiteren Informationen dargestellt werden. Hier wird neben dem Modell selbst auch eine Übersicht der Anlagen dargestellt, in der der Experte bereits verknüpfte wie auch unverknüpfte Anlagen aufgelistet sieht und in der Auflistung suchen kann. Diese unverknüpften Anlagen können mit einer Komponente aus dem 3D-Modell verknüpft werden, um eine örtliche Zugehörigkeit der Anlage zu zeigen. Der Status einer verknüpften Anlage wird im 3D-Modell farblich dargestellt. Zudem werden nach Auswahl einer Anlage weitere Details zum DT wie Stammdaten und aktuelle Messwerte angezeigt, damit der Experte schon in dieser Ansicht für ihn relevante Daten finden kann, ohne erst die entsprechende Anlagenübersicht aufrufen zu müssen (Abbildung 5). Für eine bessere Übersicht wird es dem Experten in der erweiterten Ansicht ermöglicht, mehrere Ebenen bzw. verschiedene 3D-Modelle anzeigen zu lassen.

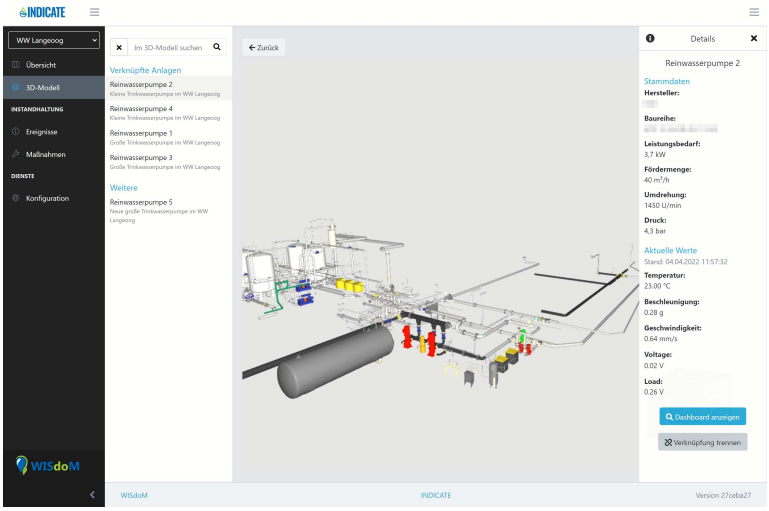


Abb. 5: 3D-Modell mit weiteren Informationen zur ausgewählten Anlage

Weiterhin spielen im Rahmen der Instandhaltung die Bereiche Maßnahmen und Ereignisse eine zentrale Rolle. Diese Bereiche dienen der Einsicht angelegter Maßnahmen sowie wichtiger Ereignissen innerhalb des Expertensystems. Die Übersichten stellen die Daten in tabellarischer Form dar, welche um ein Textfeld zur Filterung ergänzt werden, damit das Instandhaltungspersonal beliebig in dieser suchen und die Maßnahmen bzw. Ereignisse auf relevante Einträge reduzieren kann. Bei den Maßnahmen gibt es die vier Maßnahmentypen Instandhaltung, Verbesserung, Wartung und Inspektion, für die jeweils ein eigenes Protokoll besteht, welches vom Instandhaltungspersonal ausgefüllt werden kann. Beim Auftreten eines Ereignisses wird dem Experten eine Benachrichtigung, unabhängig von der aufgerufenen Seite, angezeigt. Zudem wird die Anzahl der ungelesenen Ereignisse im Navigationselement der Ereignisse dargestellt.

Für den Bereich Dienste, welcher für den Experten für PdM entwickelt wird, werden in einer Komponente zunächst die konfigurierbaren Dienste aufgelistet, in welcher der Experte auf einen Blick erkennen kann, welche Dienste zurzeit aktiv sind, und in der er weitere (Meta-) Informationen, wie u. a. zugeordnete Anlagen bzw. Schnittstellen, finden kann. Die weitere Komponente dient der Verwaltung eines Dienstes, in der der Experte die Konfiguration eins Dienstes anpassen kann und so bspw. andere Sensoren hinterlegen, Modelle für die Anomalieerkennung auswählen oder andere Grenzwerte festlegen kann.

Für die Umsetzung einer strukturieren Dateiverwaltung wird ein Datei-Explorer implementiert, mit Funktionen zum Dateiupload bzw. -download. Zur Strukturierung der Dateien werden diese in Ordner verschoben werden, welche ebenfalls erstellt und gelöscht werden können. Das Dateisystem ist nicht als eigenständige Seite über die Navigation erreichbar, sondern in der Stationsübersicht und der Anlagenübersicht eingebettet.

3 Diskussion und Ausblick

Der implementierte Prototyp zeigt die Möglichkeit, Komponenten eines Wasserwerks als DT abzubilden. Durch die Verwendung des BIM-Modells des Wasserwerk Langeoogs, welches durch den Projektpartner zur Verfügung gestellt wurde, konnte ein digitales Abbild der Pumpen webbasiert im Kontext des Wasserwerks entwickelt werden. Hierdurch kann die Anwendung das gesamte Wasserwerk digital betrachten und erhält durch die Integration von Sensordaten sowie Stammdaten und aktuelle Messwerte bereits einen großen Mehrwert des Modells. In Verbindung mit den an den Reinwasserpumpen angebrachten Sensoren konnte auch der automatisierte Live-Datenaustausch vom Wasserwerk zum Expertensystem umgesetzt werden. Hierbei hat sich gezeigt, dass eine konstante Verbindung zu den Sensoren benötigt wird, um Informationen jederzeit live darstellen zu können und für den Anwender einen Mehrwert bieten zu können. Dennoch gelten hier folgende Limitationen: Weder durften die Pumpen angesteuert werden, noch ist ein Aktuator zur Steuerung verfügbar. Dies ist dadurch begründet, dass das Wasserwerk als kritische Infrastruktur eingestuft wurde, und eine Steuerung auf Seiten des Expertensystems daher nicht möglich war. Das Expertensystem unterstützt einen bidirektionalen Datenaustausch, welcher in der Definition eines DT maßgeblich ist. Für den fehlenden Aktuator wurde ein Ein- und Ausschaltmechanismus für Anlagen als Exemplarische Komponente entwickelt, welcher das Verhalten imitiert. nach einer kritischen Betrachtung erfüllt das Expertensystem nicht vollständig die Anforderungen eines DT, mindestens jedoch die Integrationstiefe des Digital Shadow. Für die vollständige Umsetzung des DTs könnten Aktuatoren angebracht werden, welche dann mittels eigenen Microservices unkompliziert integriert werden können. Weitere technische Anlagen des Wasswerk wurden nicht abgebildet, jedoch wird die Grundlage im Expertensystem geboten, sodass beliebig viele Anlagen eines anderen Typs abgebildet werden können.

Da sich hier auf das Wasserwerk in Langeoog fokussiert wurde, muss bzgl. der zweiten Leitfrage zunächst erörtert werden, welche Instandhaltungsstrategie aktuell zum Einsatz kommt. In dem Wasserwerk wird die präventive Instandhaltung angewandt, mit jährlichen Inspektionen und täglichen manuellen Inspektionen sowie einer Generalüberholung, die alle fünf Jahre stattfindet. Bei diesen Instandhaltungsstrategien kann es vorkommen, dass eine Inspektion oder Instandhaltung durchgeführt wird, obwohl diese gar nicht benötigt wird, was u. U. dazu führt, dass Anlagen bzw. Ersatzteile bereits ausgetauscht werden, bevor deren Abnutzungsvorrat ausgeschöpft ist. Um eine optimale Nutzung zu erreichen, eignen sich sowohl PdM als auch Smart Maintenance, da beide Strategien den Experten frühzeitige Informationen über die Anlagen geben und so Wartungen nur nach Bedarf durchgeführt werden können. Als PdM-Methoden können physikalische, modellbasierte, datengetriebene oder hybride Varianten genutzt werden.

Das Expertensystem nutzt Methoden des Machine Learnings, um aus den gesammelten Daten Modelle zur Prognose von Messwerten zu erstellen. Inwiefern die Modelle einen langfristigen Zustandsverlauf durch die Messwerte und somit potenziellen Verschleiß der

Pumpe abbilden können, lässt sich nicht fundiert evaluieren, da hierfür einerseits Expertenwissen über die Zustandsüberwachung der Pumpen erforderlich ist, andererseits sich der Pumpenverschleiß erst über mehrere Jahre abbildet und für die Prognose eine ausreichende Datenhistorie und Referenzdaten zu wartungsbedürftigen Pumpen bzw. Pumpenausfällen notwendig sind. Prognosen der Modelle sollten somit von Experten interpretiert werden, um daraus Maßnahmen abzuleiten. Weiterhin ist das Expertensystem in der Lage, Anomalien aus den gewonnenen Datensätzen der Sensorik zu erkennen. Diese können den Experten weitere nützliche Informationen über den Zustand einer Anlage geben. Um Anomalien zu erkennen, müssen die Experten mit Domänenwissen die Dienste im Expertensystem entsprechend konfigurieren und Grenzen definieren. Um den Gesundheitszustand umfassender abzubilden, könnte es zukünftig interessant sein, weitere Messgrößen zu betrachten, wie den Wasserdurchfluss und -druck, die Viskosität und die Temperatur des Öls oder die Frequenzumrichterdaten zur Analyse von Drehzahl und Leistung.

Hinsichtlich der dritten Leitfrage zeigt das Expertensystem, dass PdM gut auf einen DT aufgebaut werden kann, welcher grundsätzlich mit vielen Daten angereichert wird, die im Zuge für PdM benötigt werden. Die Kombination der beiden Konzepte wird mittels Dashboards und Diagrammen umgesetzt. Für jeden Sensor kann in der Anlage eine neue Schnittstelle hinzugefügt werden, um die Sensordaten dem DT zuzuführen und in Form von Diagrammen zu visualisieren. Durch die Verknüpfung mit weiteren Schnittstellen, welche die Daten aus anderen Diensten beziehen können, kann der DT auch um Prognosen und eine Anomalieerkennung erweitert werden. Für auftretende Anomalien kann konfiguriert werden, dass automatisch Ereignisse erstellt und dem Experten durch Benachrichtigungen erkenntlich gemacht werden. Von den zahlreichen, zentral vorhandenen Daten eines DTs kann die Instandhaltungsstrategie PdM profitieren, da Ausfallzeitpunkte möglichst genau prognostiziert werden sollen, um den Abnutzungsvorrat optimal auszunutzen. Weiterhin hat das Instandhaltungspersonal die Möglichkeit, Maßnahmen für einzelne Anlagen zu erstellen, welche vorausschauend schon mit Daten des DTs vorbelegt sein können.

Die Leitfragen haben während der Projektzeit dazu beigetragen, die Konzepte PdM und DT mit dem Praxispartner anhand eines konkreten Anwendungsfalls näher zu untersuchen. Schließlich konnte ein Prototyp umgesetzt werden, der die Kombination der beiden Konzepte aufweist. Allerdings konnte die Instandhaltungsstrategie PdM nicht dahingehend entwickelt werden, für Anlagen möglichst genaue Ausfallzeitpunkte vorherzusagen oder anhand der vorhandenen Sensordaten bestimmte Störungsmuster erkennen zu können. Die Datenlage gab eine solche Verarbeitung nicht her und die Sensor-Schnittstelle zum Abfragen von Live-Schwingungsdaten zum Durchführen einer Schwingungsanalyse funktionierte nicht erwartungsgemäß. Der Fehler seitens des Anbieters wurde erst zum Projektende behoben und konnte nicht weiter verfolgt werden. Zusammengefasst gab es die Herausforderung der fehlenden Daten, den Winterbetrieb der Anlagen im Wasserwerk und Verbindungsabbrüche im Mobilfunk zum Wasserwerk wegen eines eingeschränkten Datenvolumens. Abschließend lässt sich festhalten, dass mit dem Expertensystem sowohl die Struktur als

auch viele Funktionalitäten geboten werden, um einen DT abbilden zu können und um das Instandhaltungspersonal bei der Durchführung von PdM unterstützen zu können.

Literaturverzeichnis

[Fu20] Fuller, Aidan; Fan, Zhong; Day, Charles; Barlow, Chris: Digital Twin: Enabling Technologies, Challenges and Open Research. IEEE Access, 8:108952–108971, 2020.

[Gü20] Güntner, Georg; Hoher, Simon; Eberle, Michael; Glachs, Dietmar; Kranzer, Simon; Schäfer, Georg; Schranz, Christoph: Digital Twins im Anlagen-Lebenszyklus. Bericht, Salzburg Research, 2020.

[Ho19] Holländer, Robert: Chancen und Herausforderungen der Verknüpfungen der Systeme in der Wasserwirtschaft (Wasser 4.0). Dessau Umweltbundesamt: Dessau-Roßlau, Germany, 2019.

[Kr18] Kritzinger, Werner; Karner, Matthias; Traar, Georg; Henjes, Jan; Sihn, Wilfried: Digital Twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 51(11):1016–1022, 2018.

[Ra19] Ran, Yongyi; Zhou, Xin; Lin, Pengfeng; Wen, Yonggang; Deng, Ruilong: A Survey of Predictive Maintenance: Systems, Purposes and Approaches. arXiv preprint arXiv:1912.07383, 2019.

[Sc20] Schmucker, Benedikt; Ellinger, Johannes; Benker, Maximilian; Semm, Thomas; Zäh, Michael F: Einsatz eines Digitalen Zwillings zur Prozessoptimierung und prädiktiven Instandhaltung. Zeitschrift für wirtschaftlichen Fabrikbetrieb, 115(s1):78–83, 2020.

[Wy21] Wybrands, Marius; Frohmann, Fabian; Andree, Marcel; Marx Gómez, Jorge: WISdoM: An Information System for Water Management. In (Kamilaris, Andreas; Wohlgemuth, Volker; Karatzas, Kostas; Athanasiadis, Ioannis N., Hrsg.): Advances and New Trends in Environmental Informatics. S. 131–146, 2021.

Entwicklung eines bedarfsgerechten Umweltinformationsportals mit Bürgerbeteiligung

Ali Akyol[1] Andreas Solsbach[2] Jorge Marx Gómez [3]

Abstract: Der vorliegende Beitrag zeigt die Ergebnisse des zweijährigen Projektes „Portal für eine bedarfsgerechte Umweltinformationsbereitstellung von Unternehmen und Kommunen unter Bürgerbeteiligung“ (Akronym: PUUK), gefördert von der Deutschen Bundesstiftung Umwelt, auf. Das anwachsende Bewusstsein von Bürgern für Umwelt erfordert zunehmende fachliche Kenntnisse zur Interpretation bereitgestellter Umweltinformationen. Umweltinformationen zum Feinstaub in der Heimatstadt der Bürger werden als PM2,5 oder PM10 zur Verfügung gestellt, wobei eine Interpretation dieser Werte ohne Fachkenntnisse über die Kennzahlen und damit verbundener Grenzwerte viele Fragen beim Bürger offen lässt. An dieser Stelle will das Projekt PUUK unterstützen, indem zunächst der Informationsbedarf der Bürger mittels einer Onlinestudie ermittelt wird und anschließend ein bedarfsgerechtes Umweltinformationsportal entwickelt wird.

Keywords: Partizipation; Umweltinformationsportal; Umweltinformationen; Bürgerpartizipation [4]

1 Steigendes Umweltbewusstsein in der Gesellschaft und Politik

Umweltinformationen bekommen durch das zunehmende Bewusstsein für Umwelt- und Klimaschutzthemen in der Gesellschaft eine steigende Bedeutung [BfU18]. Ein Beispiel für das steigende Bewusstsein der Gesellschaft zeigt die Umwelt- und Klimaschützerin Greta Thunberg mit der Schülerbewegung „Fridays for Future“, die im Jahr 2018 in Schweden ins Leben gerufen wurde [KN19]. Laut der Europäischen Umweltagentur sterben in der EU mehr als 400.000 Menschen und weltweit mehr als 700.000 jährlich an den Folgen von Luftverschmutzung. In Deutschland fordert die Luftverschmutzung jährlich ca. 70.000 Todesfälle [Ta20, Gi20, Eu20, De21].

Umweltinformationen werden dabei als Daten definiert, die sich mit vielen verschiedenen Aspekten der Umwelt beschäftigen. Dazu gehören beispielsweise Informationen über Luft, Wasser, Boden, oder anderen natürlichen Lebensräumen. Auch zu Umweltinformationen gehören Indikatoren wie Lärm, Strahlung oder Energie. Umweltinformationen verfügen in

[1] Universität Oldenburg, Department für Informatik, Ammerländer Heerstr. 114-118, 26129 Oldenburg, ali.akyol@uol.de, https://orcid.org/0000-0002-0682-1948

[2] Universität Oldenburg, Department für Informatik, Ammerländer Heerstr. 114-118, 26129 Oldenburg, andreas.solsbach@uol.de, https://orcid.org/0000-0002-8799-0478

[3] Universität Oldenburg, Department für Informatik, Ammerländer Heerstr. 114-118, 26129 Oldenburg, jorge.marx.gomez@uol.de, https://orcid.org/0000-0002-8799-0478

[4] Aufgrund der besseren Lesbarkeit wird im Text das generische Maskulinum verwendet. Gemeint sind jedoch immer alle Geschlechter

doi:10.18420/inf2022_71

der Regel über einen zeitlichen und räumlichen Bezug [BfU15]. Mit den Geoinformationen und dem zeitlichen Bezug werden statistische Auswertungen und die Erkennung von Trends in der Entwicklung von umweltrelevanten Indikatoren ermöglicht.

Seit der Einführung einer verpflichtenden nichtfinanziellen Berichterstattung vom Deutschen Bundestag sind Unternehmen und indirekt Kommunen aufgefordert, Umweltinformationen vorzuhalten, um den stetig steigenden Anforderungen von Politik und Öffentlichkeit gerecht zu werden (Gesetz zur Stärkung der nichtfinanziellen Berichterstattung der Unternehmen in ihren Lage-und Konzernlageberichten (CSR-Richtlinie-Umsetzungsgesetz)) [Bu17].

In welcher Form und Art Bürgern Umweltinformationen bereitgestellt werden sollen ist jedoch weitgehend unbekannt, da wenig Wissen über den Informationsbedarf besteht. Im Portal der Stadt Berlin findet sich die Projektidee „Umweltinformation - die Ampel für alle Bürger“ zu einer IT-Lösung, die Bürger bei der Deckung des Informationsbedarfes unterstützen soll. Diese Idee aus Dezember 2021 zeigt, dass aktuell noch immer Handlungsbedarf besteht und dass die damaligen Anforderungen aus 2019 an das Projekt PUUK noch immer relevant sind [Er21].

Die Problematik des unbekannten Informationsbedarfes führt dazu, dass Unternehmen und Kommunen nicht wissen oder nur schwer abschätzen können, ob die bereitgestellten Informationen den Anforderungen der Bürger gerecht werden. Ein weiteres Problem stellt der Prozess der Bereitstellung der Umweltinformationen dar. Die Bereitstellung von Informationen nach Anfragen zu Umweltinformationen werden bisher individuell durchgeführt, sodass die Unternehmen und Kommunen Personalressourcen dafür bereitstellen müssen. Hinzu kommt die Herausforderung, dass Umweltinformationen für Laien oftmals nur schwer verständlich sind. Das macht weitere Beschreibungen nötig, um die Informationen interpretieren zu können. Es existieren Barrieren zum Zugang zu Umweltinformationen. Diese werden im Kapitel 2.3 näher erläutert. Im Rahmen dieses Forschungsprojektes wurde ein Umweltinformationsportal entwickelt, das vielseitig und von unterschiedlichen Stakeholdern verwendet werden kann. Die Key-User von PUUK sind Unternehmen, Kommunen und Bürger, die jeweils eine tragende Rolle in dem Umweltinformationsportal einnehmen.

2 Ermittlung der aktuellen Bedarfe an Umweltinformationen

Zur Ermittlung der aktuellen Bedarfe wird eine Onlinestudie mit über 660 Teilnehmern durchgeführt. Die Studie ist in fünf Fragengruppen unterteilt und hat zwei zentrale Fragestellungen.

2.1 Ermittlung der Bürgerbedarfe durch eine Onlinestudie

Onlinestudien sind seit den 1990er Jahren ein bewährtes Instrument zur Datenerhebung in der Markt- und Sozialforschung. Sie können mit geringem Aufwand umgesetzt werden

und die Auswertungen der Ergebnisse können digital und schnell erfolgen [BC15]. Die durchgeführte Onlinestudie wurde so konzipiert, dass die Teilnahme anonym war. Es wurden keine verpflichtenden Angaben erfragt, mit denen es möglich wäre, die Teilnehmer anhand der Antwort zu identifizieren. Bei der Frage nach der Postleitzahl wurden nur die ersten vier Ziffern erfragt, um den Radius des Wohnortes zu erhöhen. Dies erschwert das Zurückverfolgen auf den einzelnen Teilnehmer. Die Onlinestudie wurde durchgeführt, da nur wenig Wissen über die Nutzer von Umweltportalen und ihren Informationsbedarf bekannt ist. Mit den Ergebnissen der Studie sollen Anforderungen definiert werden, die als Funktionen im Umweltinformationsportal implementiert werden können. Die Kernziele der Onlinestudie:

- Welche Nutzergruppen besuchen Umweltinformationsportale?
- Was sollten die Schwerpunkte eines Umweltinformationsportals sein?

Die durchgeführte Onlineumfrage besteht aus 31 Fragen, die in fünf Kategorien aufgeteilt wurden. Die Umfrage wurde anonym durchgeführt, jedoch hatten die Teilnehmer die Möglichkeit, ihre E-Mail zu hinterlegen, sodass sie über Neuigkeiten zum Forschungsprojekt PUUK informiert werden können. Entwickelt wurde der Fragebogen von folgenden PUUK Projektpartnern: Universität Oldenburg, leanact, Stadt Oldenburg und Klimaschutz- und Energieagentur Niedersachsen.

Nr.	Fragengruppe
1.	Zu Ihrer Person
2.	Allgemeines Interesse an Umweltdaten
3.	Für Sie relevante Umweltinformationen
4.	Von Ihnen genutzte Umweltinformationsportale
5.	Ein für Sie entwickeltes Umweltinformationsportal sollte folgende Aspekte erfüllen

Tab. 1: Die Fragengruppen der Onlinestudie [PP21].

Der Fragebogen wurde so aufgebaut, dass zunächst allgemeine Fragen zu dem Teilnehmer gestellt wurden, um zu identifizieren um welche Art von Bürger es sich bei dem Teilnehmer handelt. In den weiteren Kategorien wurden anschließend Fragen zum Interesse des Teilnehmers an Umweltinformationen, den bereits genutzten Umweltinformationsportalen und aus Sicht des Teilnehmers notwendigen Aspekten eines Umweltinformationsportals gestellt, um mögliche Funktionalitäten aus den gewünschten Aspekten abzuleiten.

2.1.1 Mittel zum Abrufen von Umweltinformationen

Bei der Beantwortung der Frage: „Mit welchem Endgerät würden Sie sich am häufigsten über Umweltinformationen erkundigen wollen?“ belegte das Smartphone den ersten Platz mit 1471 Punkten. Den zweiten Platz belegten Computer/Laptops mit 1437 Punkten. Den dritten Platz belegte das Tablet mit 496 Punkten vor „Sonstiges“, mit 151 Punkten. Die Teilnehmer

konnten dabei ihre Prioritäten festlegen, mit welchem Endgerät sie sich am ehesten über Umweltinformationen erkundigen würden. Hierfür wurde folgende Bewertungsskala angewendet: Priorität (1): 3 Punkte, Priorität (2): 2 Punkte, Priorität (3): 1 Punkt und Priorität (4): 0 Punkte. Das Smartphone wurde 312 Mal mit Priorität (1), 243 Mal mit Priorität (2) und 49 Mal mit Priorität (3) eingestuft. Hieraus ergibt sich für das Smartphone folgende Bewertung: (312 x 3) + (243 x 2) + (49) = 1471 Punkte.

2.1.2 Quellen der Umweltinformationen

Bei dieser Frage wurden die Teilnehmer gefragt, wie sie sich über Umweltinformationen erkundigen. Dabei gaben 82% der Befragten an, dass sie sich mithilfe von Internetseiten über Umweltinformationen informieren. Die zweithäufigsten Stimmen hat die Option „soziale Netzwerke“ erhalten, die genau 50% der Befragten darstellt. Die restlichen Antwortmöglichkeiten werden durch traditionelle Medien repräsentiert. „Fernsehen“ und „Zeitung“ halten mit einer Differenz von 35 bzw. 99 Stimmen zu sozialen Netzwerken.

Internetseiten	Soziale Netzwerke	Fernsehen	Zeitung	Radio	Sonstiges
537	328	295	229	125	76

Tab. 2: Quellen für Umweltinformationen nach Häufigkeit [PP21] (Mehrfachantworten möglich).

2.1.3 Aufschlüsselung der sozialen Netzwerke

Falls die Teilnehmer der Onlinestudie „soziale Netzwerke“ als Möglichkeit angaben, um sich über Umweltinformationen zu informieren, wurde eine zusätzliche Frage nach dem genutzten sozialen Netzwerk aktiviert. Dabei konnten die Teilnehmer zwischen verschiedenen sozialen Netzwerken wählen. Auf Platz eins ist mit großem Abstand Instagram mit 249 Stimmen benannt worden (Facebook: 110; Whatsapp und Twitter folgten mit ähnlichen Anteil auf Platz drei und vier).

Instagram	Facebook	WhatsApp	Twitter	LinkedIn	Telegram	Tiktok	Sonstiges
249	110	71	69	29	14	7	22

Tab. 3: Quellen für Umweltinformationen nach Häufigkeit [PP21] (absolute Zahlen, Mehrfachantworten möglich).

2.2 Zusammenfassung der Onlinestudie

Es konnten Erkenntnisse darüber gewonnen werden, welche Begriffe in einem Umweltportal definiert und beschrieben werden sollten. Außerdem gaben 537 Teilnehmer an, dass sie Umweltinformationen von Internetseiten beziehen und 328 Personen nutzen Soziale Netzwerke als Quelle für ihre Umweltinformationen. Zudem konnten Ergebnisse über die

Häufigkeit der Besuche eines Umweltportals gewonnen werden. 66 Teilnehmer gaben an, dass sie Umweltportale täglich besuchen. Insgesamt gaben 187 Teilnehmer an, wöchentlich Umweltportale zu besuchen. Auch ist zu erwähnen, dass 222 Teilnehmer angaben, Umweltportale gar nicht zu besuchen. Mehr als die Hälfte von den 222 Teilnehmer sagen, dass sie kein bisheriges Wissen über Umweltinformationsportale haben (Barriere Nummer 1 in Kapitel 2.3). Als zweit häufigster Grund wurde fehlende Zeit angegeben [PP21].

2.3 Barrieren zum Zugang zu Umweltinformationen

Durch die durchgeführten qualitativen sowie quantitativen Forschungsmethoden (Onlinestudien und Workshops), konnten neue Erkenntnisse zum Zugang zu Umweltinformationen gewonnen werden. Kernergebnisse der Studien sind unter anderem, dass Barrieren zum Zugang zu Umweltinformationen bestehen. Unter die quantitative Forschungsmethode fällt die durchgeführte Onlinestudie (siehe Kapitel 2.1). Die Erkenntnisse der Onlinestudie wurden in Workshops aus verschieden Perspektiven vertieft. Als qualitative Forschungsmethode wurden aufeinander aufbauende Workshops mit unterschiedlichen Bürgern durchgeführt. Am ersten Workshop konnten alle Altersgruppen von Bürgern teilnehmen. Anschließend wurde ein Workshop mit Studierenden sowie zwei Workshops mit Schülern aus zwei Gymnasien in Oldenburg durchgeführt, um insbesondere Erkenntnisse von Generationen zu erlangen, die durch zukünftige Umweltauswirkungen stärker belastet werden.

Nach der Auswertung der Ergebnisse konnten fünf Barrieren zu Umweltinformationen identifiziert werden.

1. **Bürger kennen Umweltinformationsportale nicht.**
   Die Onlinestudie (siehe Kapitel 2.1) ermittelt, dass die meisten Bürger Umweltportale nicht oder nur wenig kennen. Zudem zeigt sie, dass sich die Mehrheit der Nutzer aus Interesse über das Wetter informiert und um ihre Aktivitäten zu planen (Wetterauskunft), jedoch nicht über die Luftqualität. Gleichzeitig gaben 61% der Teilnehmer an, dass sie Umweltportale gar nicht oder nur monatlich benutzen. Dabei gaben 83% der Teilnehmer an, dass sie Umweltportale nicht benutzen, weil sie kein Wissen über Umweltinformationsportale haben, keine Zeit oder kein sie ansprechendes Umweltportal kennen.
2. **Bürger verstehen die zur Verfügung gestellten Inhalte nicht.**
   Umweltinformationen sind für Laien nicht einfach zu verstehen, sodass sie darauf angewiesen sind, sich über externe Quellen weiter zu informieren. Dies stellt eine weitere Hürde für die Benutzung von Umweltportalen dar. Zudem fehlt dem Laien das Wissen zur Interpretation der bereitgestellten Umweltinformationen, um die gesundheitlichen Auswirkungen abschätzen zu können.
3. **Bürger müssen sich selbst nach Grenzwerten erkundigen.**
   Push-Benachrichtigungen sind in den meisten Umweltportalen nicht verfügbar, sodass

die Bürger sich proaktiv über die Umweltinformationen erkundigen müssen. Das führt dazu, dass die Bürger Grenzwertüberschreitungen nicht oder nur im Nachhinein erfahren. Auch die Erstellung von Benutzerprofilen ist nicht möglich, so dass keine Personalisierung erfolgen kann.

4. **Umweltinformationen werden in den Alltag nicht integriert.**
Die bestehenden Umweltportale ermöglichen es den Bürgern nicht, stets über ihre natürliche Umwelt informiert zu sein. Der Zugang zu Umweltinformationen ist zeitaufwendig und nicht einfach gestaltet, sodass wichtige Informationen nicht im Alltag der Bürger integriert sind. Die von PUUK durchgeführte Onlinestudie und durchgeführten Workshops haben ergeben, dass die Integration von Umweltinformationen in den Alltag erwünscht ist. Beispiele sind Unwetterwarnungen und Prognosen von Staus, sodass die Nutzer rechtzeitig eine alternative Route fahren können.

5. **Umweltinformationen werden nicht gemessen.**
Dies ist eine der größten und häufigsten Barrieren, die es zu überwinden gilt, da an vielen Orten in Städten nicht gemessen wird. Bürger haben durch diese Barriere keine Möglichkeit, sich über ihre natürliche Umwelt zu erkundigen. Diese Barriere kann nur mit hohem Aufwand beseitigt werden, indem (eigene) Sensoren beschafft bzw. erworben und aufgebaut werden.

3 Ist-Zustand der Umweltportale in Deutschland

Zur Darstellung des Ist-Zustands in Deutschland wurde eine Onlinerecherche und Identifikation der Umweltportale der jeweiligen Bundesländer durchgeführt. Alle Umweltportale wurden nacheinander bezüglich ihrer Funktionalitäten analysiert. Das Umweltportal mit den meisten aufrufen zu identifizieren scheiterte daran, dass die Domains der Umweltinformationsportale meistens auf Subdomains eingerichtet sind. Die verwendeten Tools konnten nur die Topleveldomains nach Aufrufen analysieren [Hy22].

Unsere Recherchen ergaben, dass keines dieser Umweltinformationsportale die Informationen individualisiert bereitstellt. Die zur Verfügung gestellten Informationen sind für alle Besucher identisch. Einige Umweltinformationsportale können Umweltinformationen über ihr Bundesland hinaus anbieten, andere wiederum nicht. Der aktuelle Stand auf dem Markt zeigt, dass keine einheitliche Lösung für alle Bürgergruppen bereitsteht. Das Abonnieren von Benachrichtigungen wie z. B. bei Unwetterwarnungen oder Grenzwertüberschreitungen von Feinstaub (PM2,5 und PM10) sind nicht möglich.

4 Konzept eines bedarfsgerechtes Umweltinformationsportals

Anhand der ermittelten Anforderungen, die durch die Onlinestudie, den Workshops und Experteninterviews entstanden sind, kann das Konzept des neuartigen Umweltinformationsportals entwickelt werden. Der Prototyp ist in unterschiedliche Bereiche aufgeteilt.

Nr.	Bundesland	URL des Portals
1.	Baden-Württemberg [Um22b]	https://umwelt-bw.de
2.	Bayern [Ba22]	https://umweltnavigator.bayern.de
3.	Berlin [Be22a]	https://berlin.de/umwelt
4.	Brandenburg [Mi22d]	https://berlin.de/umwelt
5.	Bremen [Di22]	https://bauumwelt.bremen.de
6.	Hamburg [Be22b]	https://hamburg.de/umwelt
7.	Hessen [Hl22, He22]	https://umwelt.hessen.de, https://hlnug.de
8.	Mecklenburg-Vorpommern [La22a]	https://uis-mv.de
9.	Niedersachsen [Ni22]	https://numis.niedersachsen.de
10.	Nordrhein-Westfalen [Mi22b]	https://umweltportal.nrw.de
11.	Rheinland-Pfalz [Mi22a]	https://portalu.rlp.de
12.	Saarland [St22]	https://saarland.de
13.	Sachsen [Ko22]	https://umwelt.sachsen.de
14.	Sachsen-Anhalt [Mi22c]	https://umwelt.sachsen-anhalt.de
15.	Schleswig-Holstein [La22b]	https://schleswig-holstein.de
16.	Thüringen [Th22]	https://umweltportal.thueringen.de

Tab. 4: Umweltinformationsportale in Deutschland (eigene Quelle).

Die Keyuser in dem entwickelten Prototyp sind die Bürger und die Unternehmen. Zur Bereitstellung von externen Umweltinformationen wurden alle verfügbaren Daten vom UBA (Umweltbundesamt) zu öffentlichen Luft- und Wassermessstationen in das Umweltportal integriert [Um22a]. Anhand von Messwerten werden dem Nutzer weitere Informationen angezeigt, sodass Laien an dieser Stelle wissen, was diese Informationen für sie bedeuten. Um die Messung für die Bürger, Unternehmen oder Kommunen ermöglichen zu können, wurden eigene Sensorboxen entwickelt.

Bürger erhalten die Möglichkeit, ein Profil zu erstellen, um ihr Dashboard gestalten zu können. Das Dashboard zeigt unterschiedliche Kacheln mit Informationen an, die von den Nutzern individuell gestaltbar sind. Neben der Karte für das aktuelle Wetter wird auch eine Wetterprognose für die nächsten Tage angezeigt. Die Nutzer haben die Möglichkeit, ihre Stadt einzugeben und die Karten ein- und ausblenden zu lassen. Die Reihenfolge der Karten ist ebenfalls modifizierbar.

Aus den Umweltdaten werden Umweltinformationen, indem neben den Messdaten Zeit- und Geoinformationen hinzugefügt werden. Erst wenn alle Daten vorliegen, werden aus den Umweltdaten Umweltinformationen gewonnen [BfU15]. Alle 20 Minuten sendet die Sensorbox die gemessenen Umweltinformationen an das PUUK Umweltportal. Zudem besteht die Möglichkeit, die Umweltinformationen auch in Echtzeit abzurufen. Außerdem überträgt die Sensorbox die gemessenen Umweltinformationen in einem Zeitintervall von 20 Minuten zum Speichern an den Server, sodass die Nutzer sich die gemessenen Werte ihrer Sensorbox auch zu einem späteren Zeitpunkt anzeigen lassen können.

Die Abbildung 3 zeigt die Architektur des Umweltinformationsportals von PUUK. Das

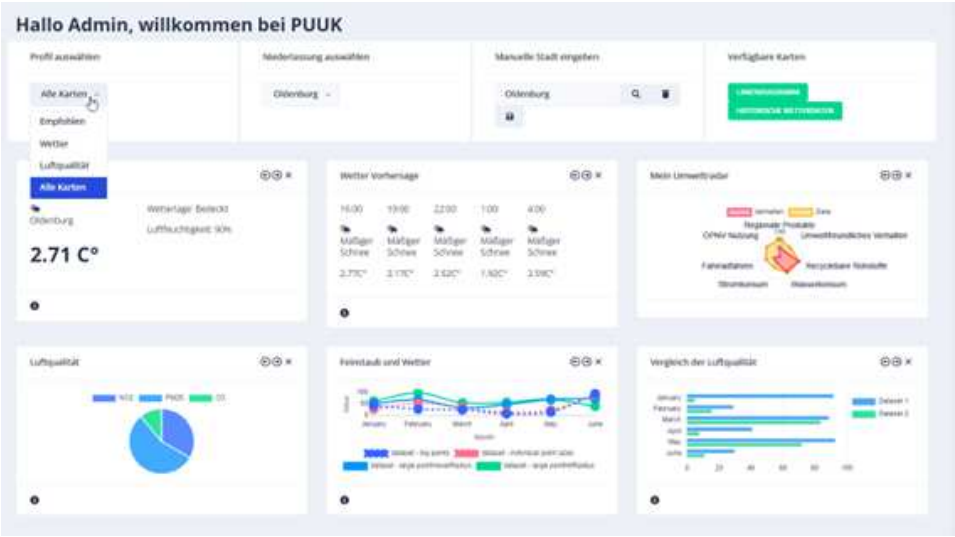


Abb. 1: Screenshot des Dashboards im entwickelten Prototyp (eigene Quelle).

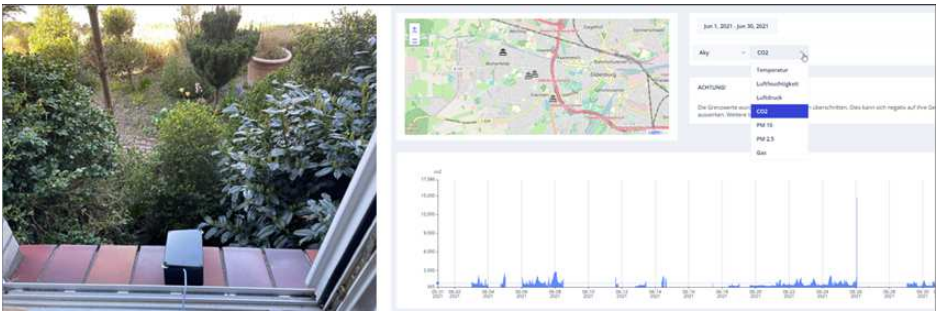


Abb. 2: PUUK Sensorbox auf der Fensterbank, welche mit dem Portal verbunden ist (eigene Quelle).

prototypisch entwickelte Umweltinformationsportal adressiert vier verschiedenen Nutzergruppen, die Umweltinformationen bedarfsgerecht abrufen können. Durch die unterschiedlichen Rollen und der Möglichkeit der Profilerstellung (Admins, Bürger, Unternehmen und Kommunen) besteht die Möglichkeit, dass Informationen personalisiert dargestellt werden können. Die Darstellung der Umweltinformationen ist auf verschiedene Endgeräte angepasst. Entwickelt wurde das Umweltinformationsportal mit dem MEAN-Stack (MongoDB, Express, Angular und Node), welches ein javascript-basiertes Framework ist [Mo22b, Mo22a, De22, Fo22, No22]. Die PUUK Sensorboxen können von allen Nutzergruppen eingesetzt werden. Mit der PUUK Sensorbox ist es möglich, dass die Nutzer Umweltinformationen über das Umweltinformationsportal in Echtzeit abrufen können.

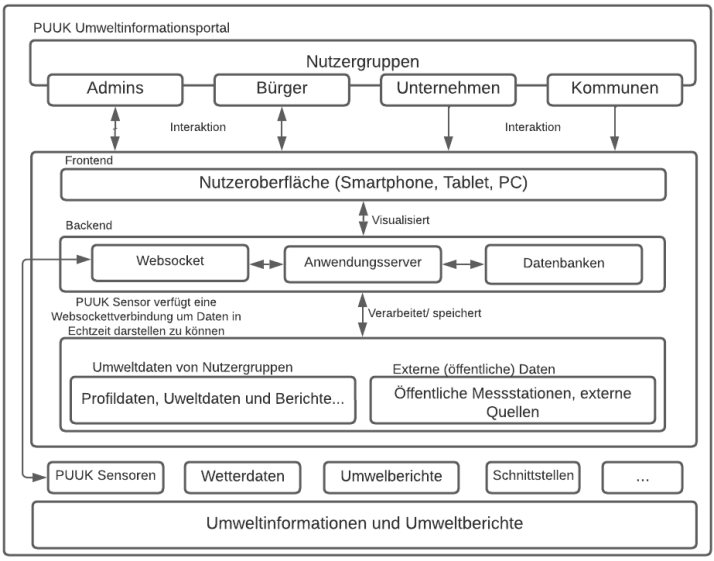


Abb. 3: PUUK Architektur (eigene Darstellung).

5 Zusammenfassung und Ausblick

Die Onlinestudie, im Rahmen des Projektes PUUK zeigt auf, dass Bürger Interesse an Umweltinformationen zeigen, jedoch eine Unkenntnis über potenzielle Informationsquellen bzw. deren Funktionalitäten besteht. Hierzu wurden im Rahmen der Analyse beispielsweise Barrieren zum Zugang ermittelt, die von aktuellen Lösungen nicht überwunden werden. Durch das Konzept der Bereitstellung eines Portals als Anlaufstelle für Bürger sollen Umweltinformationen an einer Stelle gebündelt werden. Die prototypische Umsetzung zeigt auf, wie per Schnittstellen diese Anforderung umgesetzt und durch zusätzliche Informationsbereitstellung Umweltinformationen erläutert werden.

Das Portal ist abhängig von der Datenqualität und -menge, die durch externe Datenquellen bereitgestellt oder erhoben wird. Beispielhaft konnten die Daten der öffentlichen Messstation der Stadt Oldenburg in das Portal PUUK integriert werden. Diese Messstation in Innenstadtnähe (am Heiligengeistwall) gibt stündlich Auskunft über die Luftqualität „in der Stadt“. Jedoch sind die Informationen kritisch zu bewerten, da die Luftqualität einer Stadt nicht mit nur einer Luftmessstation allein bewertet werden kann. Eine höhere Granularität der Luftqualität in den verschiedenen Stadtteilen wäre notwendig, um beispielsweise Radfahrern eine möglichst feinstaubarme Fahrradroute vorzuschlagen. Eine mögliche Lösung wäre, die Innenstadt mit weiteren Sensoren auszustatten, damit die Ergebnisse der Sensoren für einen kleineren Bereich gelten. Das Umweltinformationsportal wurde gemeinsam mit den Sensorboxen prototypisch umgesetzt. Die nächsten Schritte sind, das Umweltinformati-

onsportal in einen marktreifen Zustand zu entwickeln, damit Kommunen es in ihrer Stadt einführen können. Eine prototypische Einführung des Umweltinformationsportals in der Stadt Oldenburg steht noch aus.

Danksagung Gefördert mit Mitteln der Deutschen Bundesstiftung Umwelt. Projekt Förderkennzeichen: 35152/01. Projekttitel: Portal für eine bedarfsgerechte Umweltinformationsbereitstellung von Unternehmen und Kommunen unter Bürgerbeteiligung

Literaturverzeichnis

[Ba22] Bayerisches Staatsministerium für Umwelt und Verbraucherschutz: Umwelt Navigator Bayern: Das Portal zu Bayerns Umweltinformationen. https://www.umweltnavigator.bayern.de/, 2022.

[BC15] Barz, Heiner; Cerci, Meral: Online-Befragung. In: Frauen in Kunst und Kultur: Zwischen neuem Selbstbewusstsein und Quotenforderungen, S. 95–106. Springer Fachmedien Wiesbaden, Wiesbaden, 2015.

[Be22a] BerlinOnline Stadtportal GmbH & Co. KG: Umweltportal - Berlin.de. https://www.berlin.de/umwelt/, 2022.

[Be22b] Betreibergesellschaft hamburg.de GmbH & Co. KG: Umweltschutz und Klimaschutz in Hamburg. https://www.hamburg.de/umwelt, 2022.

[BfU15] Bundesministerium für Umwelt, Naturschutz und nukleare Sicherheit: Umweltinformationen. https://www.bmuv.de/themen/bildung-beteiligung/umweltinformation/umweltinformation-worum-geht-es#c31345, 09.07.2015.

[BfU18] Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz: Umweltinformation. https://www.bmu.de/themen/bildung-beteiligung/umweltinformation/, 2018.

[Bu17] Bundestag, Deutscher: Gesetz zur Stärkung der nichtfinanziellen Berichterstattung der Unternehmen in ihren Lage- und Konzernlageberichten (CSR-Richtlinie-Umsetzungsgesetz). S. 802–814, 2017.

[De21] Deutsche Welle: Coronavirus: Feinstaub verschlimmert COVID-19-Risiko | DW | 17.11.2020. https://www.dw.com/de/coronavirus-feinstaub-verschlimmert-covid-19-risiko/a-55628089, 15.05.2021.

[De22] Developers, Google: Angular - What is Angular? https://angular.io/guide/what-is-angular, 09.07.2022.

[Di22] Die Senatorin für Klimaschutz, Umwelt, Mobilität, Stadtentwicklung und Wohnungsbau Bremen: Klimaschutz, Umwelt, Mobilität, Stadtentwicklung und Wohnungsbau Bremen. https://www.bauumwelt.bremen.de/, 2022.

[Er21] Erlach, Sandra: Umweltinformation - die Ampel für alle Bürger. https://mein.berlin.de/ideas/2021-13442/, 2021.

[Eu20] Europäische Umweltagentur: Vorzeitige Todesfälle durch Luftverschmutzung. https://www.eea.europa.eu/de/pressroom/newsreleases/zahlreiche-europaeer-sind-immer-noch/vorzeitige-todesfaelle-durch-luftverschmutzung, 2020.

[Fo22] Foundation, OpenJS: Express - Node.js-Framework von Webanwendungen. https://expressjs.com/de/, 10.07.2022.

[Gi20] Giegold, Sven: Mehr als 400.000 Tote durch Luftverschmutzung: EU-Kommission muss entschlossener handeln. https://sven-giegold.de/luftverschmutzung-kommission-handeln/, 23.11.2020.

[He22] Hessisches Landesamt für Naturschutz, Umwelt und Geologie: Umweltmesswerte in Hessen | Hessisches Landesamt für Naturschutz, Umwelt und Geologie. https://umwelt.hessen.de/, 2022.

[Hl22] Hlnug: Umweltmesswerte in Hessen | Hessisches Landesamt für Naturschutz, Umwelt und Geologie. https://www.hlnug.de/messwerte, 21.05.2022.

[Hy22] HypeStat: Web Statistics and Analysis | HypeStat. https://hypestat.com/, 2022.

[KN19] Koos, Sebastian; Naumann, Elias: Vom Klimastreik zur Klimapolitik: Die gesellschaftliche Unterstützung der „Fridays for Future “-Bewegung und ihrer Ziele: Forschungsbericht. 2019.

[Ko22] Kommunikation und Öffentlichkeitsarbeit, Referat: Startseite - Umwelt - sachsen.de. https://www.umwelt.sachsen.de/, 2022.

[La22a] Landesamt für Umwelt, Naturschutz und Geologie Mecklenburg-Vorpommern: Umweltinformationssystem Mecklemburg-Vorpommern. https://www.uis-mv.de/, 2022.

[La22b] Landesportal Schleswig-Holstein: Webseite des Bundeslandes Schleswig-Holstein. https://www.schleswig-holstein.de/DE/Home/home_node.html, 2022.

[Mi22a] Ministerium für Klimaschutz, Umwelt, Energie und Mobilität: PortalU - Umweltportal Rheinland-Pfalz. https://www.portalu.rlp.de/portal/;jsessionid=A8A60ADABB8DA5081EC7AC97088D31B2, 2022.

[Mi22b] Ministerium für Umwelt, Landwirtschaft, Natur- und Verbraucherschutz des Landes Nordrhein-Westfalen: Umweltportal NRW - Startseite - Umweltportal NRW. https://www.umweltportal.nrw.de/, 2022.

[Mi22c] Ministerium für Wissenschaft, Energie, Klimaschutz und Umwelt: Umweltportal Sachsen-Anhalt. https://www.umweltdaten.brandenburg.de/, 2022.

[Mi22d] Ministerium für Wissenschaft, Energie, Klimaschutz und Umwelt: Landwirtschafts- und Umweltinformationssystem Brandenburg (LUIS-BB). https://www.umweltdaten.brandenburg.de/, 21.05.2022.

[Mo22a] MongoDB: MongoDB: The Developer Data Platform | MongoDB. https://www.mongodb.com/, 11.07.2022.

[Mo22b] MongoDB: What Is The MEAN Stack? Introduction & Examples. https://www.mongodb.com/mean-stack, 11.07.2022.

[Ni22] Niedersächsisches Ministerium für Umwelt: NUMIS - Das niedersächsische Umweltportal. https://numis.niedersachsen.de/portal/, 2022.

[No22] Node.js: Node.js. https://nodejs.org/en/, 09.07.2022.

[PP21] PUUK-Projekt: Ergebnisse der Onlinestudie. https://puuk-projekt.de/wp-content/uploads/2021/03/PUUK-Projekt-Umfrage-Zwischenauswertung-18.02.21.pdf, 20.01.2021.

[St22] Staatskanzlei des Saarlandes: Webseite Saarland. https://www.saarland.de/DE/home/home_node.html, 2022.

[Ta20] Tagesschau: EU-Bericht: 400.000 Tote jährlich durch dreckige Luft. https://www.tagesschau.de/ausland/europa/eureport-umweltverschmutzung-101.html, 08.09.2020.

[Th22] Thüringer Ministerium für Umwelt, Energie und Naturschutz: Umweltportal Thueringen. https://www.umweltportal.thueringen.de/, 2022.

[Um22a] Umweltbundesamt: Messstationen des Umweltbundesamtes. https://www.umweltbundesamt.de/daten/luft/luftdaten/stationen/, 24.03.2022.

[Um22b] Umweltministerium Baden Wüttenberg: Portal: Umwelt-BW. https://www.umwelt-bw.de/, 2022.

Low-Code Basierte Entwicklung eines Chatbots für Betriebliches Compliance Management im Umweltbereich

Heiko Thimm [1] und Vanessa Schmidt[2]

Abstract: Die hohe Regelungsdichte des Umweltrechts und die permanente Zunahme neuer Vorschriften stellt Unternehmen vor große Herausforderungen bei der Aufrechterhaltung der Umweltrechtskonformität. Neue Vorschriften müssen auf ihre Relevanz für die Unternehmensstandorte bewertet werden, was ohne geeignete IT-Unterstützung eine aufwändige und fehleranfällig Tätigkeit darstellt. In einem Forschungsprojekt wird an der Entwicklung von intelligenten Assistenzsystemen für diese Kernaufgabe des Compliance Managements im Umweltbereich gearbeitet. Der Beitrag gibt einen Überblick über ein Cloud-basiertes Chatbot Assistenzsystem, das derzeit auf Basis der weit verbreiteten Plattform Dialogflow entwickelt wird.

Keywords: Umweltrecht, Compliance Management, Assistenzsystem, Chatbot

1 Einleitung

Ein wichtiges Teilgebiet des betrieblichen Umweltmanagements ist die Überwachung und Durchsetzung der Umweltrechtskonformität, die im Englischen als ‚Environmental Compliance Management' (ECM) bezeichnet wird. Dabei muss durch entsprechende Überwachungs-, Bewertungs-, Prüf-, Entscheidungs- und Dokumentationsaufgaben sichergestellt werden, dass das Unternehmen an allen Standorten und für alle Produkte die jeweils geltenden relevanten Gesetze, Verordnungen und Bestimmungen (nachfolgend subsumiert unter dem Begriff Vorschriften) fristgerecht einhält. Vorschriften können dabei von unterschiedlichen Regelsetzern stammen wie kommunale Umweltbehörden, Landesbehörden, das Umweltbundesamt und andere Institutionen auf europäischer, internationaler und supra-nationaler Ebene. IT Lösungen zur Unterstützung und/oder (Teil-)Automatisierung dieser Aufgaben bieten unternehmerische Chancen wie zum Beispiel Kosteneinsparungen, mehr Sicherheit bei der Einhaltung von Erledigungsfristen und die Entlastung von Experten durch die Übertragung von Teilaufgaben an Personen, die nicht über entsprechende Expertenkenntnisse verfügen. Im Kundenservice, Reklamationsmanagement, Wartungs- und Instandhaltungsmanagement und anderen betrieblichen Bereichen hat sich der Einsatz von Chatbot Technologien zunehmend etabliert. Im Gegensatz dazu finden Chatbots im Zusammenhang mit operativen Umweltmanagementaufgaben bisher sowohl in der Praxis als auch in der Fachliteratur nur

[1] Hochschule Pforzheim, Fakultät für Technik, Tiefenbronner Str. 65, 75175 Pforzheim, heiko.thimm@hs-pforzheim.de, https://orcid.org/0000-0000-0000-0000
[2] Hochschule Pforzheim, Fakultät für Technik, Tiefenbronner Str. 65, 75175 Pforzheim, vanessa.schmidt@hs-pforzheim.de

doi:10.18420/inf2022_72

relativ wenig Beachtung. Diese Situation war Anlass für die Entwicklung eines Chatbot Prototyps, der von betrieblichen Compliance Managern als Cloud Anwendung genutzt werden kann. Der Prototyp wird auf Basis von Dialogflow [Mu20], [SA20] entwickelt, einer kommerziellen Low-Code Chatbot Entwicklungsumgebung von Google. In die Entwicklung des Prototyps flossen Ergebnisse früherer Forschungsprojekte ein, wie ein generisches Prozessmodell für ECM [Th15], ein Entscheidungsbaumverfahren für Relevanzbewertungen [Th21] und Abstraktionen zur Modellierung von unternehmensspezifischen ECM Gegebenheiten [Th22]. Im vorliegenden Beitrag wird ein allgemeiner Überblick über dieses laufende Projekt und die bisher gewonnenen Erkenntnisse gegeben. Der Beitrag kann nach Ansicht der Autoren wertvolle Hilfestellungen und auch Inspiration für weitere Chatbot Entwicklungen im Kontext des betrieblichen Umweltmanagements liefern.

Im nachfolgenden Kapitel 2 werden wesentliche Aspekte der Relevanzbewertungsaufgabe beschrieben gefolgt von einem kurzen Technologieüberblick über Chatbots in Kapitel 3. Kapitel 4 gibt einen Überblick über den Prototyp und erste Erkenntnisse aus den bisherigen Entwicklungsarbeiten. Kapitel 5 enthält eine Zusammenfassung und einen Ausblick auf zukünftige Forschungsarbeiten.

2 Relevanzbewertung von Vorschriften als ECM Kernaufgabe

Die ECM Aufgabenstellungen sind vielfältig und erfordern umfassende Expertenkenntnisse. Von zentraler Bedeutung ist dabei ein operativer Geschäftsprozess, der vier Kernaufgaben umfasst [Th15]: 1. Überwachung von Ankündigungen neuer Vorschriften bzw. Revisionen von Vorschriften durch entsprechende Regelsetzer, 2. Relevanzbewertung neuer Vorschriften und Revisionen für alle betroffenen Unternehmensstandorte, 3. Entscheidung über Maßnahmen zur Einhaltung der geänderten Vorschriftensituation, 4. Wirksamkeitskontrollen von Maßnahmen. Die Erledigung dieser Aufgaben muss stringent dokumentiert werden, denn nur beim Vorliegen einer lückenlosen und manipulationssicheren Dokumentation sämtlicher umweltrechtlicher Belange können Umweltaudits bestanden werden. Um der Nachweis- und Dokumentationspflicht nachzukommen, wird deshalb häufig ein zentrales d.h. unternehmensweites elektronisches Rechtskataster geführt. Zur Durchführung von Relevanzbewertungen sind sowohl umfassende Kenntnisse im Umweltrecht (z.B. Bodenschutz, Emissionsschutz, Gewässerschutz, Chemikalienmanagement) als auch detaillierte Kenntnisse über Produkte, Herstellungsverfahren, die technische Infrastruktur und sonstige Gegebenheiten an den Standorten zwingend erforderlich. Sinnvollerweise wird im Rahmen von Relevanzbewertungen auch die Notwendigkeit für Maßnahmen betrachtet. Unter anderem aus diesem Grund geht es bei Relevanzbewertungen nicht nur darum, ob eine neue Vorschrift/Revision relevant ist oder nicht relevant ist, sondern auch darum, ob sie in der Zukunft relevant werden könnte und inwiefern Maßnahmen erforderlich sind.

Für mittelgroße global agierende Unternehmen mit Produktionsstandorten in verschiedenen Weltregionen gehen ECM Praktiker davon aus, dass monatlich mehrere Duzend neue in Englisch und anderen Sprachen verfasste Vorschriften und/oder Änderungen bestehender Vorschriften zu prüfen sind. Es ergeben sich daraus im Schnitt vier bis sechs neue Verpflichtungen, wovon durchschnittlich zwei bis drei Verpflichtungen Maßnahmen erfordern. Oftmals ist es insbesondere für kleinere Unternehmen schwierig das erforderliche Zeitbudget für qualifizierte und erfahrene ECM Spezialisten vorzuhalten. In Engpasssituationen wird deshalb teilweise auf externe Service Dienstleister zurückgegriffen. Im vorliegenden Projekt wird u.a. untersucht, inwieweit diese betriebliche Problematik reduziert werden kann, indem Mitarbeiter bei der Durchführung von Relevanzbewertungen von einem Chatbot-basierten Assistenzsystem unterstützt werden, das als Cloud Lösung zur Verfügung steht. Das Assistenzsystem ist insbesondere konzipiert für Mitarbeiter, die über keinen idealtypischen Kenntnis- und Erfahrungshintergrund im Bereich des Compliance Managements verfügen. Sie sollen vom System in die Lage versetzt werden, typische Relevanzbewertungsaufgaben eigenständig durchzuführen. Nur noch in besonders schwierigen Bewertungssituationen sollen die ECM Experten eingebunden werden. Besonders für Unternehmen mit mehreren Standorten kann ein solches Cloud-basiertes Assistenzsystem viele Vorteile bieten.

3 Chatbot Technologie

Ein Chatbot ist ein Dialogsystem, das eine vollautomatisierte text- und/oder sprachbasierte Kommunikation mit dem Nutzer ohne direkten menschlichen Eingriff bereitstellen kann. Neben dem Einsatz von Chatbots im betrieblichen Umfeld finden Chatbots zunehmend auch in der öffentlichen Verwaltung Einsatz. Einer europäischen Studie zu Folge sollen sie zukünftig die heute üblichen Kataloge mit elektronischen Diensten ablösen, um Bürgern und anderen Nutzergruppen eine höhere Servicequalität zu bieten [Eu19].

Benutzer können über Chat-Plattformen wie Slack, Alexa, WhatsApp oder auch über E-Mail mit dem Chatbot kommunizieren. Dabei handelt es sich um eine Kommunikation zwischen Menschen und einer Maschine, die einfache Anfragen nach dem Wetter oder komplexere Anfragen wie Kundenberatungen übernehmen kann. Mithilfe einer Künstlichen Intelligenz (KI) kann auf nicht vorhersehbare Fragen in natürlicher Sprache geantwortet werden. Auch können Bilder, unstrukturierte Daten und Emotionen erkannt und verarbeitet werden. Kann eine Anfrage von der Maschine nicht verstanden oder beantwortet werden, wird in der Praxis oft der weitere Dialog von einem menschlichen Bearbeiter übernommen.

3.1 Arten von Chatbots

In der Fachliteratur (z.B. [AM20]) werden Chatbots häufig in drei verschiedene Arten eingeteilt, die nachfolgend kurz vorgestellt werden sollen. *Regelbasierte Chatbots* nutzen

Techniken des Natural Language Processing (NLP) und des Knowledge Representation (KR). Bei regelbasierten Chatbots werden Befehle, Regeln und Keywords vordefiniert, die für einen strukturieren Gesprächsablauf verwendet werden. Diese Art von Chatbot lässt sich vor allem bei einfachen Gesprächsverläufen einsetzen. Hierbei wird mit vordefinierten potentiellen Fragestellungen und Schlüsselwörtern herausgefunden, welche Absicht der Nutzer hat und mit einem passenden vordefinierten Pfad reagiert. Die Nutzerabsichten werden in verschiedenen Chatbot Frameworks auch als `Intents´ bezeichnet. Bei den Schlüsselwörtern zur Intent Erkennung handelt es sich um Begriffe, die Entitäten bezeichnen, also individuelle physische oder abstrakte Objekte der realen Welt. *KI-basierte Chatbots* besitzen ‚intelligente' Fähigkeiten. Hierzu zählt vor allem das Verstehen von Kontexten durch Einsatz von Technologien des Natural Language Understanding (NLU) und das selbständige Erlernen von Fähigkeiten aus strukturierten und unstrukturierten Daten. Besondere Herausforderungen sind unter anderem das Erkennen von Missverständnissen und Rechtschreibfehlern sowie das Erlernen von Dialekten und Emotionen. Dazu werden entsprechende Trainingsdaten benötigt. *Hybride Chatbots* basieren auf einer Kombination von Techniken, die bei regelbasierten und bei KI-basierten Chatbots zum Einsatz kommen. Hybride Chatbots handeln in erster Linie auf Basis von regelbasierten vordefinierten Keywords und Prozessen. Die KI Komponente dient dazu, Interaktionen mit dem Nutzer zu analysieren, um Informationen über den Nutzer wie z.B. das Interessensfeld, Empfehlungen für den Chatverlauf sowie Trainingsdaten abzuleiten. Wird die Konversation an einen Mitarbeiter weitergeleitet, generiert der Chatbot Lösungsvorschläge für den Live-Chat mit dem Nutzer.

3.2 Grundkomponenten von Chatbots

Im Allgemeinen bestehen Chatbots aus vier Grundkomponenten, die ggf. durch eine optionale Komponente für Analysezwecke ergänzt werden [Sc22]. In Abbildung 1 sind diese fünf Komponenten dargestellt, auf die nachfolgend kurz eingegangen werden soll.

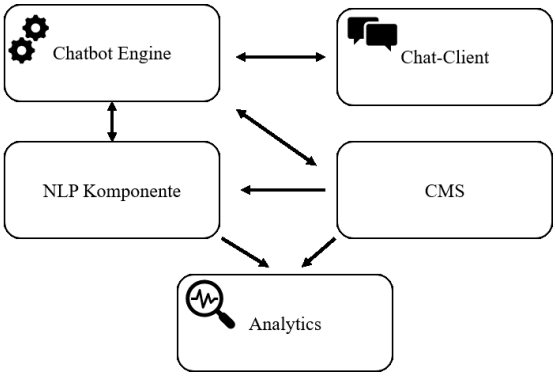


Abb. 1: Grundkomponenten eines Chatbots in Anlehnung an [Sc22]

Der Chat-Client nimmt text- oder sprachbasierte Nachrichten des Nutzers entgegen, leitet diese an die Chatbot Engine (Server) als Anfragen weiter, nimmt die Antworten des Servers entgegen und gibt diese aus [Eu19]. Viele vorhandene Chat-Tools können als Chat-Client verwendet werden wie zum Beispiel Facebook-Messenger, WhatsApp oder Slack. Die Chatbot Engine verwaltet (ggf. parallele) Sitzungen mit Chat-Clients und verarbeitet deren Anfragen. Sitzungskontexte (z.B. Chatbot Aktionen) werden gespeichert im Content Management System (CMS), um darin vorhandene Informationen für den weiteren Dialogverlauf zu nutzen. Wenn ein Nutzerdialog in einen Live-Chat mit einem Mitarbeiter überführt werden muss, übernimmt die Chatbot Engine dabei die Koordination und die Steuerung. Die NLP Komponente wird zur Verarbeitung von Freitexteingaben des Nutzers und zur Extraktion von Informationen aus dem bisherigen Sitzungskontext benötigt [Eu19]. Mit den Ergebnissen wird versucht, eine vorliegende Freitexteingabe einer im System angelegten passenden Fragestellungen zuzuordnen. Sollte keine Fragestellung mit einer vorgegebenen Mindestwahrscheinlichkeit zugeordnet werden können, werden durch Auswertung des Sitzungskontexts weitere Information vom Nutzer eingeholt. Wenn die Fragestellung erkannt ist, wird darauf mit einer hinterlegten Nachricht geantwortet. Von Domänenexperten sind potentielle Fragestellungen der Nutzerzielgruppe und Antwortmöglichkeiten zu erstellen. Beides wird in der Literatur oft als Chatbot Content bezeichnet. Je umfassender und passgenauer der Content auf den Auskunftsbedarf der Nutzer abgestimmt ist, desto gewinnbringender der Einsatz des Chatbots für das Unternehmen und desto höher die Nutzerakzeptanz. Der Content wird im CMS System gespeichert. Bei KI-basierten Chatbots werden auch Trainingsdaten im CMS verwaltet. Manche Chatbot Engines verfügen über ein integriertes Analytics Modul, um statistische Auswertungen zur Nutzung des Chatbots zu ermöglichen. Die Auswertungen geben wichtige Hinweise zur Verbesserung des Contents und zur Anpassung an geänderte Benutzeranforderungen und Verhaltensweisen des Benutzers. Besonders hilfreich sind Analysen zur Aufrufhäufigkeit. Sie geben Aufschluss darüber, wie oft und wofür welche Nutzer den Chat geöffnet und in welchem Zustand sie ihn wieder verlassen haben. Statistische Auswertungen zu den vordefinierten Intents sind ebenfalls eine wichtige Grundlage zur Optimierung und Weiterentwicklung des Chatbot Contents.

3.3 Dialogflow

Für die Entwicklung von Chatbots gibt es spezielle Entwicklungsframeworks und Bibliotheken für verschiedene Programmiersprachen wie z.B. die Bibliothek NLTK für Python. Darüber hinaus sind auch Plattformen vorhanden, die einen Low-Code Entwicklungsansatz ermöglichen wie zum Beispiel die Ontologie-basierte Dialogplattform ODP S3 von paragon semvox [Pa17] und die Plattform Dialogflow von Google [Mu20], [LL19], [SA20], mit der Chatbots und Softwarelösungen für den Google Assistant als Cloud Anwendung erstellt werden können. Das Architekturbild in Abbildung 2 zeigt den logischen Aufbau eines auf Dialogflow-basierenden Chatbots. Im Kontext von Dialogflow werden potentielle Nutzeranfragen als Intents bezeichnet, die als parametrisierbare Datenobjekte im System angelegt werden. Bei der Zuordnung von

Intents, erfragt das System so lange Informationen vom Nutzer, bis sämtliche Parameterwerte zur weiteren Verarbeitung zur Verfügung stehen. Dieser Vorgang wird bei Dialogflow als Slot Füllung bezeichnet [SA20]. Standardmäßig reagiert das System auf einen zugeordneten Intent mit einer vordefinierten (statischen) Antwort. Es besteht jedoch die Möglichkeit durch Verwendung der Integrations- und Schnittstellendienste dynamische Antworten für Intents anzulegen. Dazu ist für den Intent die Auftragsausführung (Fulfillment) zu aktivieren. Dies bewirkt, dass zur Laufzeit eine Antwort generiert wird, die durch Aufruf eines definierten Dienstes (Webhook request) zustande kommt. Dieses flexible Schnittstellenkonzept ermöglicht für die Antwortgenerierung externe Datenquellen, externe Dienste, Webanwendungen und Apps zu nutzen.

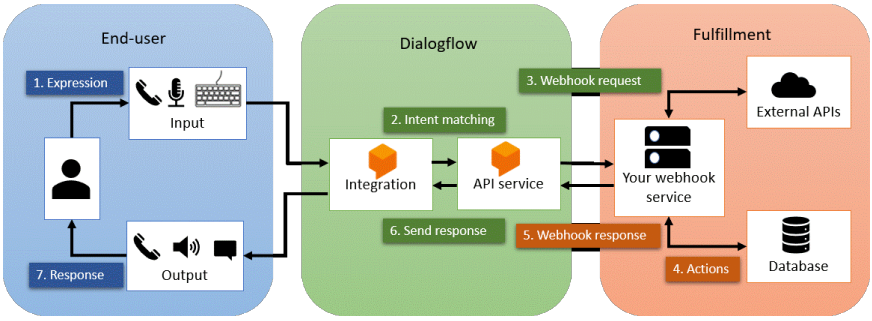


Abb. 2: Systemarchitektur Dialogflow in Anlehnung an [Go21]

4 Prototyp

Um die Möglichkeiten und Grenzen eines auf Dialogflow basierenden Low-Code Entwicklungsansatzes für den intendierten Chatbot praktisch auszuloten, wurde in einem ersten Schritt ein einfacher Demonstrator eines regelbasierten Chatbots entwickelt [Sc21]. Der Demonstrator kann zur Abfrage von Informationen verwendet werden, die zur Durchführung von Compliance Management Aufgaben und Entscheidungen benötigt werden. Da der Demonstrator darüber hinaus keine eigenständigen Aktionen ausführen kann, handelt es sich bei diesem System nach der Klassifikation von Steil und Wrede [SW19] um einen Informationsassistenten. Die gewonnen Erfahrungen mit Dialogflow werden in dem in diesem Artikel vorgestellten Projekt für die Entwicklung eines Cloud-basierten Assistenzsystems genutzt. Das System soll regelbasiert zu eigenständigen Aktionen fähig sein und dadurch Relevanzbewertungen umfassend unterstützen. Die Verwendung von Dialogflow birgt jedoch auch die üblichen Nachteile eines Low-Code Entwicklungsansatzes. Die Modellierungskonstrukte und strukturellen Implementierungsvorgaben des zu Grunde liegenden Ansatzes von Dialogflow bringen zwangsläufig ggf. enge Freiheitsgrade für die Problemabbildung sowie technische Implementierungsbeschränkungen mit sich. Sollte sich bereits bei der laufenden Implementierung oder in der Testphase der verfolgte Low-Code Ansatz daher nicht bewähren, ist geplant, das Assistenzsystem in der Programmiersprache Python zu implementieren.

Im nachfolgenden wird das zurzeit auf Basis von Dialogflow entwickelte Assistenzsystem näher beschrieben. Das System wird mit Echtdaten (z.B. Rechtskataster, Standortdaten) eines an mehreren deutschen Standorten für die Automotiv und Möbel Branche mit rund 6000 Mitarbeitern produzierenden Unternehmens aufgebaut. In mehreren Stufen soll gemeinsam mit den betrieblichen Umweltmanagern des Partnerunternehmens die Gebrauchstauglichkeit des Systems getestet werden.

4.1 Chatbot Funktionsprinzip, Sitzungsschema und Content

Bei vielen Chatbot Anwendungsszenarien, wie zum Beispiel bei Kunden Self-Service Szenarien, muss zunächst das Anliegen des Nutzers durch entsprechende Fragetechniken herausgefunden werden. Beim intendierten Chatbot ist das generelle Benutzeranliegen jedoch bereits im Voraus bekannt. Es kann deshalb direkt mit einer auf eine Relevanzbewertung zugeschnittene und vom System geführte Sitzung begonnen werden. Ausgehend von bisherigen Forschungsergebnissen wurde unter Mitwirkung des Praxispartners das in Tabelle 1 dargestellte generische Grundschema einer typischen Dialogsitzung mit dem Chatbot entworfen. Das Schema implementiert eine vom Chatbot angeleitete explorative Analyse der Vorschrift bzw. Revision und generiert für jeden zu berücksichtigenden Unternehmensstandort einen eigenen Bewertungsvorschlag durch Auswahl einer der folgenden Alternativen: 1. relevant und sofortige Maßnahmen erforderlich, 2. relevant und Maßnahmen in naher Zukunft erforderlich, 3. relevant und keine Maßnahmen erforderlich, 4. aktuell nicht relevant aber möglicherweise zukünftig relevant, 5. momentan und auch in Zukunft nicht relevant.

Der Benutzer hat die Möglichkeit, Bewertungsvorschläge anzunehmen und diese vom Chatbot in die entsprechenden Rechtkataster eintragen zu lassen. In den Schritten des Schemas werden unterschiedliche Bestandteile des unternehmensspezifischen Contents wissensbasiert verarbeitet. Die domänenspezifische Wissensbasis besteht aus Logik-basierten Regeln und aus Heuristiken sowie einem umfassenden Entitätskatalog. Der Entitätskatalog beinhaltet eine strukturierte Sammlung von Schlüsselwörtern, die in Vorschriften zur Beschreibung von regulatorischen Bezügen und regulatorischen Vorgaben verwendet werden. Schlüsselwörter, die im 4. Schritt dazu dienen den regulatorischen Bezug einer Vorschrift zu bestimmen, beschreiben zum Beispiel Substanzen in Produkten, im Abwasser und in der Abluft. Außerdem sind Schlüsselwörter vorhanden, die Substanzen beschreiben, die in Herstellungsverfahren als indirektes Material benötigt werden (z.B. Kühlmittel bei spanenden Verfahren in der Metallverarbeitung) oder in Herstellungsprozessen neu entstehen. Beispiele für Schlüsselwörter zur Erkennung von regulatorischen Vorgaben (5. Schritt) sind Begriffe zur Beschreibung von Grenzwerten, Verboten von Inhaltsstoffen, Gefahrstoffe, geforderte Schutzvorrichtungen oder Schutzmaßnahmen.

Der unternehmensspezifische Content umfasst allgemeine Unternehmensinformationen und für jeden relevanten Standort ein Rechtskataster und ein Compliance Management Profil. In diesem Profil ist u.a. beschrieben, welche Bereiche des Umweltrechts für den

Standort von Belang sind. Außerdem sind im Profil sämtliche für den Standort relevanten regulatorischen Aspekte beschrieben, auf die regelmäßig in Vorschriften Bezug genommen wird. Diese Informationen werden insbesondere im 6. Schritt benötigt, um zu ermitteln, für welche Unternehmensstandorte die Vorschrift bzw. Revision keinerlei Relevanz hat und deshalb von der weiteren Untersuchung ausgeschlossen werden können.

Nr.	Inhalt/Ziel	Kommentar
1	Bestimmung relevanter Metadaten der Vorschrift	Interaktionsschritt m. Benutzereingaben
2	Bestimmung des territorialen Geltungsbereichs der Vorschrift	Interaktionsschritt m. Benutzereingaben
3	Bestimmung des/der Umweltrechtgebiets/e	Interaktionsschritt m. Benutzereingaben
4	Bestimmung des/der regulatorischen Bezugs/Bezüge	Interaktionsschritt m. Benutzerauswahl aus Entitätskatalog
5	Bestimmung der regulatorischen Vorgabe/n	Interaktionsschritt m. Benutzerauswahl aus Entitätskatalog
6	Ermittlung der Unternehmensstandorte, die von der Vorschrift betroffen sein könnten	Systemschritt
7	Ausgabe Ergebnisliste u. Auswahl eines Standorts über den weitere Infos abgerufen werden können	Interaktionsschritt m. Benutzerauswahl aus Ergebnisliste oder zurück zu einem früheren Schritt
8	Abfrage ob regulatorische Vorgabe(n) am Standort eingehalten wird/werden	Interaktionsschritt m. Benutzereingaben
9	Abfrage ob Notwendigkeit für Maßnahmen besteht	Interaktionsschritt m. Benutzereingaben
10	Ermittlung eines standortspezifischen Bewertungsvorschlags	Systemschritt
11	Ausgabe Bewertungsvorschlag u. Auswahl zur weiteren Verarbeitung	Interaktionsschritt m. Benutzerauswahl: Vorschlag annehmen, Vorschlag verwerfen, zurück zu einem früheren Schritt
12	Ausgabe Auswahloptionen u. Auswahl zur weiteren Verarbeitung	Interaktionsschritt m. Benutzerauswahl: angenommene Vorschläge in

		Rechtkataster übertragen, Vorschläge, verwerfen, zurück zu einem früheren Schritt, Sitzung beenden

Tab. 1: Grundschema eines Chatbot Dialogs zur Unterstützung von Relevanzbewertungen

4.2 Implementierungsüberblick und erste Erkenntnisse

Es wurde mit der Implementierung der in Abbildung 3 gezeigten logischen Systemarchitektur begonnen. Der aktuelle Fokus liegt dabei auf der Abbildung des oben beschriebenen Sitzungsschemas auf die von Dialogflow vorgesehenen Konzepte. Dabei sollen auch die NLP/NLU Funktionen von Dialogflow zum Einsatz kommen. Mehrere Alternativen befinden sich derzeit in der Entwicklung. Bei den Alternativen handelt es sich um verschiedene Abbildungsmöglichkeiten des Sitzungsgrundschemas. Die bisherigen Abbildungsalternativen bestehen aus bis zu fünf unterschiedlich parametrisierten Intents. Die Anzahl der statischen und dynamischen Antworten, die den Intents zugeordnet sind, variiert bei den Alternativen ebenfalls. Es ist davon auszugehen, dass nicht alle statischen Antworten vollständig auf die von Dialogflow bereit gestellten Funktionen abgebildet werden können. Deshalb sind in der Architektur zur Kompensierung dieser funktionellen Lücken entsprechende anwendungsspezifische Funktionen konzeptionell vorgesehen. Die dynamischen Antworten werden generiert, indem extern verwaltete unternehmensspezifische Daten (Unternehmensprofil sowie standortspezifische Rechtskataster und Compliance Management Profile) und der Entitätskatalog ausgewertet werden. Die Anbindung der unternehmensspezifischen Daten, die als Cloud-basierte Tabellenkalkulationsdokumente angelegt sind, erfolgt mittels Google Sheets unter Verwendung der SheetsDB API. Der in seiner initialen Version ebenfalls als Tabellenkalkulationsdokument vorliegende Entitätskatalog ist auf die gleiche Weise angebunden.

Wenn dynamische Antworten komplexe Prüfungen und Auswertungen von Nutzereingaben und/oder unternehmensspezifische Daten erfordern, wird dies von einer externen App durchgeführt, dem Compliance Management (CM) Agent. Dazu kann der CM Agent direkt auf den Entitätskatalog und die unternehmensspezifischen Daten zugreifen. Als Implementierungssprache für den CM Agent wird die Programmiersprache Python verwendet. Zur Anbindung an Dialogflow als Webhook Service kann zum Beispiel das Flask Framework verwendet werden.

Zur Bereitstellung eines initialen Entitätskatalogs mit einer ausreichenden Anzahl an Schlüsselwörtern, wurde eine Vielzahl deutscher Gesetzestexte zum Gewässerschutz und Emissionsschutz ausgewertet. Für diese Auswertung wurde die Text Mining Software Wordstat [De19] verwendet. Der initiale Katalog soll unter Berücksichtigung der Ergebnisse der Testphase später erweitert und in eine Datenbank-basierte Lösung überführt werden. Es ist jedoch davon auszugehen, dass das System trotzdem nicht für alle zukünftigen Vorschriften und Revisionen passende Auswahlmöglichkeiten für Terme zur

Beschreibung des regulatorischen Bezugs und/oder der regulatorischen Vorgabe/n enthalten wird. Ein „Erlernen“ passender neuer Terme durch entsprechende interaktive Eingabemöglichkeiten in den Schritten 4 und 5 des Sitzungsschemas wird daher als Möglichkeit einer Benutzer-gesteuerten bedarfsorientierten Katalogerweiterung in Erwägung gezogen. Zusätzlich sollen im Projekt auch Ansätze für ein (teil)automatisiertes „Erlernen“ relevanter neuer Terme aus Gesetzestexten und deren Ergänzung im Entitätskatalog entwickelt werden. Eine Alternative stellt dabei die Entwicklung einer eigenständigen Softwarelösung z.B. auf Basis der Programmiersprache Python und dessen Bibliotheken für Textanalyseverfahren dar. Zukünftige Versionen von Dialogflow werden über integrierte Machine Learning Funktionen verfügen und eröffnen damit eine zweite Alternative. Es bleibt jedoch abzuwarten, ob sich diese Funktionen zur Erweiterung des Prototyps um ein selbstständiges Erlernen von Termen eignet.

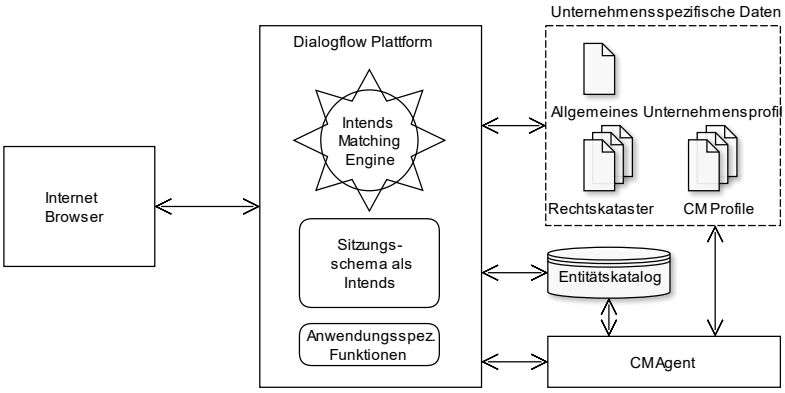


Abb. 3: Logische Architektur des Prototyps

Aus den bisherigen Implementierungsaktivitäten haben sich bisher nur wenige nennenswerte Erkenntnisse ergeben. Ob sich ein auf Dialogflow basierender Low Code Ansatz tatsächlich zur Realisierung des intendierten Cloud-basierten Assistenzsystem eignet, kann zum jetzigen Zeitpunkt noch nicht beantwortet werden. Allerdings kann man Dialogflow durchaus eine gewisse vorrangige Ausrichtung auf klassische Chatbot Anwendungen im Bereich Kundenservice und Reklamationsmanagement attestieren. Während bei diesen Anwendungen gewöhnlich mit unbekannten Nutzern kommuniziert wird, spielen die Nutzerdaten bzw. zugehörigen Unternehmensdaten bei der intendierten Anwendung eine herausragende Rolle. Das zentrale Verarbeitungsprinzip von Dialogflow ist dem allgemeinen Prinzip von Rule Engines sehr ähnlich. Bei der Aktivierung eines Intents wird aus den vorgegebenen Reaktionen die Reaktion ausgewählt und ausgelöst, die am besten zum gegebenen Kontext passt. Eine Möglichkeit zum Austausch von Parametern und/oder Variablen zwischen verschiedenen Intents ist in Dialogflow jedoch nicht vorgesehen [KL20]. Aus dem oben skizzierten Feinkonzept zur Abbildung des Sitzungsgrundschemas lässt sich erkennen, dass sich das zentrale Verarbeitungsprinzip mit dieser Einschränkung nicht besonders gut zur Implementierung komplexerer

Datenverarbeitungsschritte eignet. Solche Schritte müssen außerhalb der Dialogflow Kernkomponenten von externen Applikationen durchgeführt werden wie z.B. dem CM Agent des Prototyps.

5 Zusammenfassung und Ausblick

Im vorgestellten Projekt wird ein Cloud-basiertes Assistenzsystem entwickelt, das Benutzer bei der Bewertung neuer umweltrechtlicher Vorschriften hinsichtlich der Relevanz für Unternehmensstandorte unterstützt. Das intendierte Assistenzsystem basiert dabei auf einem Low-Code Chatbot Ansatz, obwohl die Merkmale des Anwendungsfalls nicht den klassischen Charakteristika von Chatbot Anwendungen entsprechen. Vom Low-Code Entwicklungsansatz erwartet sich das Projektteam eine frühe Einbeziehung zukünftiger Anwender, eine zielorientierte Akquisition und Aufbereitung des Domänenwissens durch die frühzeitige Einbindung der Anwender, eine kurze Entwicklungszeit und damit einen zeitnahen Proof-of-Concept. Die Konzeptprüfung wird die Frage beantworten, ob mit dem gewählten Ansatz ein gebrauchstaugliches ECM Assistenzsystem entwickelt werden kann oder ob es dazu anderer Alternativen bedarf. Andere Alternativen sind z.B. die Entwicklung eines Chatbots auf Basis der Programmiersprache Python und die Entwicklung eines klassischen wissensbasierten Assistenzsystems. Zu den Forschungsthemen, die in einer späteren Projektphase verfolgt werden, gehört der Einsatz von Textanalyseverfahren zur automatisierten Inhaltserschließung von Vorschriften und zur automatisieren Erkennung von Daten, für die der Prototyp Nutzereingaben benötigt. Die langfristige Vision des Forschungsprojekts stellt ein hybrides Chatbot Assistenzsystem dar. Vom System sollen fortlaufend gängigen Ankündigungskanäle von gesetzlichen Bestimmungen überwacht werden. Bei neuen Ankündigungen soll das System für jedes hinterlegte Kundenunternehmen aktiv werden. Für jedes Unternehmen soll im ersten Schritt eine unternehmensspezifische Relevanzbewertungen durchführt werden. Im zweiten Schritt sollen Vorschläge für ggf. erforderliche Compliance Maßnahmen generiert werden. Ein Hauptproblem stellt dabei die große Heterogenität des Dokumentenkorpus der deutschen und der europäischen Umweltgesetzgebung dar. Es fehlt den Dokumenten vor allem an Einheitlichkeit bezüglich Struktur, Form und Metadaten. Ein weiteres Feld zukünftiger Forschungsaktivitäten ist die Einbeziehung von Digital Twin Ansätzen und Technologien. Dabei ist zu untersuchen inwieweit aus verfügbaren Digital Twins von Unternehmensstandorten Daten für das Compliance Management Profil generiert werden können bzw. wie praktikable Integrationskonzepte aussehen können.

Literaturverzeichnis

[AM20] Adamopoulou, E.; Moussiades, L.: An Overview of Chatbot Technology. In (Maglogiannis, I.; Iliadis, L.; Pimenidis, E. Eds.): Artificial Intelligence Applications and Innovations. Springer International Publishing, Cham, pp. 373–383, 2020.

[De19] Deng, Q. et al.: Inside the Black Box of Dictionary Building for Text Analytics: A Design Science Approach. Journal of International Technology and Information 3/27, pp. 119–159, 2019.

[Eu19] European Commission, Directorate-General for Informatics, ISA2 Programme: Architecture for public service chatbots, 2019.

[Go21] Google: Dialogflow ES-Grundlagen. https://cloud.google.com/dialogflow/es/docs/basics?hl=de, accessed 11 Apr 2022.

[KL20] Kihlborg, M.; Lilja, A.: Important Criteria when Choosing a Conversational AI Platform for Enterprises. KTH Report TRITA-EECS-EX-2020:441, Stockholm, Sverige, 2020.

[LL19] Luber, S.; Litzel, N.: Was ist Dialogflow? https://www.bigdata-insider.de/was-ist-dialogflow-a-788455/, accessed 29 Mar 2022.

[Mu20] Muhammad, A. F. et al.: Developing English Conversation Chatbot Using Dialogflow: International Electronics Symposium IES. IEEE, pp. 468–475, 2020.

[Pa17] Paragon Semvox: Paragon Semvox ODP S3. Die führende Technologie für proaktive Interaktions- und Assistenzsysteme. http://www.semvox.de/wp-content/uploads/2018/10/paragon_semvox_ODPS3_Produktblatt_DE_2017.pdf, accessed 29 Mar 2022.

[SA20] Sabharwal, N.; Agrawal, A.: Introduction to Google Dialogflow. In (Sabharwal, N.; Agrawal, A. Eds.): Cognitive Virtual Assistants Using Google Dialogflow. Apress, Berkeley, CA, pp. 13–54, 2020.

[Sc21] Schmidt, V.: Entwicklung eines Chatbot-Prototyps für die Relevanzbewertung von Compliance-Vorgaben im Kontext des betrieblichen Umweltmanagements. Bachelor Thesis, Pforzheim, 2021.

[Sc22] Schönholzer, P.: Technischer Aufbau von Chatbots. https://botfabrik.ch/blog/technischer-aufbau-von-chatbots/, accessed 29 Mar 2022.

[SW19] Steil, J.; Wrede, S.: Maschinelles Lernen und lernende Assistenzsysteme. Berufsbildung in Wissenschaft und Praxis (BWP) 3/2019, pp. 14–18, 2019.

[Th15] Thimm, H.: IT-Supported Assurance of Environmental Law Compliance in Small and Medium Sized Enterprises. Int. Journal of Computer and Information Technology 2/4, 297-, 2015.

[Th21] Thimm, H.: Towards Decision Tree Based Assistance of a Cloud Platform for Environmental Compliance Management. In (Kamilaris, A. et al. Eds.): Adjunct Proc. 34th. Conf. Environmental Informatics, Nicosia, Cyprus, Sept. 2020. Shaker, Düren, pp. 75–83, 2021.

[Th22] Thimm, H.: Systems theory-based abstractions and decision schemes for corporate environmental compliance management. Sustainable Operations and Computers Vol. 3, pp. 188–202, 2022.

Wirtschaftsinformatik an Hochschulen Angewandter Wissenschaften

Wirtschaftsinformatik an Hochschulen Angewandter Wissenschaften

Christian Czarnecki[1], Vera G. Meister[2] und Verena Majuntke[3]

Keywords: Wirtschaftsinformatik, Hochschulen Angewandter Wissenschaften, anwendungsorientierte Forschung

Der Workshop *Wirtschaftsinformatik* (WI) *an Hochschulen Angewandter Wissenschaften* (HAW) wird in Kooperation mit dem *Arbeitskreis WI an HAW* (AKWI) organisiert. Der AKWI ist der Dachverband der Fachbereiche mit deutschsprachigen WI-Studiengängen und/oder -Studienschwerpunkten an HAW. Er versteht sich als fachkompetenter und hochschulpolitischer Ansprechpartner zu Themen des Studiums der WI an HAW sowie der WI als anwendungsbezogene Wissenschaft. Der AKWI ist als Arbeitskreis des Fachbereichstags Informatik und zugleich als Fachgruppe WI-AKWI der Gesellschaft für Informatik (GI) organisiert (siehe www.akwi.de).

Der AKWI organisiert jährlich eine Jahrestagung, die aus einem wissenschaftlichen und hochschulpolitischen Teil besteht. Die diesjährige 35. Jahrestagung fand vom 12.-13.09.2022 in Berlin statt. Für den wissenschaftlichen Teil wurden insgesamt 26 Forschungsbeiträge eingereicht, aus denen in einem doppelblinden Begutachtungsverfahren die vorliegenden vier Artikel für den Workshop auf der INFORMATIK 2022 empfohlen wurden. Alle Beiträge wurden bereits auf der AKWI-Jahrestagung präsentiert und werden im vorliegenden Workshop in überarbeiteter und erweiterter Form vorgestellt.

Die Beiträge zeigen die inhaltliche und methodische Breite der WI-Forschung an HAW. Mithilfe von Text-Mining-Methoden zur Informationsextraktion werden Schlüsselwörter im Forschungsgebiet Quantum Computing identifiziert und visualisiert. Wie Machine Learning im Innovationsmanagement eingesetzt werden kann, wird anhand einer Literaturanalyse untersucht. Als gestaltungsorientierter Ansatz wird Machine Learning zur Routenplanung unter Berücksichtigung der Emission eingesetzt und anhand eines Prototypen demonstriert. Darüber hinaus wird die Konzeption einer virtuellen Lern-App erläutert sowie deren Umsetzung und Einsatz in einer Lehrveranstaltung diskutiert.

[1] FH Aachen, Eupener Str. 70, 52066 Aachen, Deutschland, czarnecki@fh-aachen.de
[2] TH Brandenburg, Magdeburger Str. 50 | 14770 Brandenburg an der Havel, vera.meister@th-brandenburg.de
[3] HTW Berlin, Treskowallee 8, 10318 Berlin, Verena.Majuntke@HTW-Berlin.de

doi:10.18420/inf2022_73

Programmkomitee

- Prof. Dr. Wolfgang Alm (HS Aschaffenburg)
- Prof. Dr. Gunnar Auth (HS Meißen)
- Prof. Dr. Thomas Barton (HS Worms)
- Prof. Dr. Christian Czarnecki (FH Aachen)
- Prof. Dr. Christian Drumm (FH Aachen)
- Prof. Dr. Sandy Eggert (HWR Berlin)
- Prof. Dr. Ingo Elsen (FH Aachen)
- Prof. Dr. Heinrich Faßbender (FH Aachen)
- Prof. Dr. Dorothee Feldmüller (HS Bochum)
- Prof. Dr. Frank Herrmann (OTH Regensburg)
- Prof. Dr. Stephan Jacobs (FH Aachen)
- Prof. Dr. Jürgen Karla (HS Niederrhein)
- Prof. Ute Klotz (HS Luzern)
- Prof. Dr. Claudia Lemke (HWR Berlin)
- Prof. Dr. Olga Levina (TH Brandenburg)
- Prof. Dr. Verena Majuntke (HTW Berlin)
- Prof. Dr. Birte Malzahn (HTW Berlin)
- Prof. Dr. Vera Meister (TH Brandenburg)
- Prof. Dr. Frank Morelli (HS Pforzheim)
- Prof. Dr. Christof Mosler (HFT Stuttgart)
- Prof. Dr. Jörg Puchan (HS München)
- Prof. Dr. Regina Polster (HS Schmalkalden)
- Prof. Dr. Kerstin Prechel (Duale Hochschule Schleswig-Holstein, Kiel)
- Prof. Dr. Margit Scholl (TH Wildau)
- Prof. Dr. Thomas Specht (HS Mannheim)
- Prof. Dr. Juliane Siegeris (HTW Berlin)
- Prof. Dr. Katharina Simbeck (HTW Berlin)
- Prof. Dr. Jessica Slamka (HS München)
- Prof. Dr. Martin Spott (HTW Berlin)
- Prof. Dr. Ingo Stengel (HS Karlsruhe)
- Prof. Dr. Jan Stöß (HS Karlsruhe)
- Prof. Dr. Arif Wider (HTW Berlin)
- Prof. Dr. Olga Willner (HTW Berlin)
- Prof. Dr. Martin Wolf (FH Aachen)

Einsatz von Machine Learning im Innovationsmanagement

– Eine Analyse des State-of-the-Art*

Daniel Szafarski und Helmut Beckmann[1]

Abstract: In volatilen Zeiten, in denen die verfügbare Datenmenge stetig steigt, nehmen Unternehmen das IM in den Fokus, um ihre Wettbewerbsfähigkeit zu sichern. Aus diesem Grund ist eine stetig steigende Anzahl an Publikationen zu verzeichnen, welche sich aktuell insbesondere mit der frühphasigen Ideengenerierung mithilfe natürlich sprachlicher Texte beschäftigen. Zur Analyse werden sowohl unüberwachte als auch überwachte Lernverfahren angewendet, wobei insbesondere Clustering-Algorithmen sowie Klassifikations- und Regressionsanalysen genutzt werden. Die präskriptive Analyse sowie die geschäftsbezogenen Innovationen sind in diesem Zusammenhang noch kaum untersucht, weswegen sich hieraus weitergehender Forschungsbedarf ableiten lässt.

Keywords: Innovationsmanagement, Machine Learning, Künstliche Intelligenz, Digitalisierung, State-of-the-Art

1 Einleitung

Durch eine volatile Umgebung, in der sich die Kundenbedürfnisse schnell verändern und die Innovationsgeschwindigkeit des Marktes steigt, rückt das Innovationsmanagement (IM) in den Fokus von Unternehmen, um ihre Wettbewerbsfähigkeit zu sichern [Ha18]. Aufgrund der enormen Relevanz von Innovationen für ein Unternehmen wird deren Entstehungsprozess bereits seit Jahrzehnten in der Forschung und Praxis stark thematisiert [Mü20]. Auch im Bereich des IM konnte die vielseitige Anwendbarkeit von Machine Learning (ML) in zahlreichen Case Studies demonstriert werden [Mü20]. Insbesondere vor dem Hintergrund stetig steigender Datenmengen können ML-Methoden nicht nur als Teil, sondern ebenfalls zur Optimierung des Prozesses genutzt werden [MG20, Mü20].

Wie die Boston Consulting Group in ihren jährlichen Umfragen aufzeigt, gehört im Jahr 2021 für zwei Drittel der Unternehmen das Thema „Innovation“ zu den wichtigsten drei strategischen Zielen. Gleichzeitig belegt die Studie, dass 80 % der B2C Unternehmen, also Unternehmen, die an Endkunden vertreiben, lediglich grundlegende Systeme zur Erhebung und Auswertung ihrer Kundendaten nutzen. Demgegenüber stehen fehlende Kompetenzen in der Auswertung und Nutzung von Daten [Bo21].

* Dieser Beitrag ist initial im Tagungsband der 35. AKWI Jahrestagung 2022 erschienen [SB22].

[1] Hochschule Heilbronn, Studiengang Wirtschaftsinformatik, Max-Planck-Str. 39, 74081 Heilbronn, {daniel.szafarski; helmut.beckmann}@hs-heilbronn.de

doi:10.18420/inf2022_74

Demzufolge werden aktuell keine beziehungsweise kaum ML-Methoden in den Unternehmen genutzt, obwohl diese wertvolle Erkenntnisse für das IM ermöglichen. In einem kürzlich publizierten Artikel von [Ha21] wird der Einsatz von Künstlicher Intelligenz (KI) für das IM untersucht und auf den zielführenden Einsatz in Organisationen hingewiesen. Einschränkend berücksichtigt werden muss, dass lediglich die ersten Prozessschritte des IM thematisiert werden und der dritte Schritt gänzlich ausgelassen wird. Die Literatur bietet zum aktuellen Zeitpunkt kaum einen Überblick über die Nutzung von ML-Verfahren für das IM, weswegen [Ro18] eine Forschungslücke in diesem Bereich aufzeigen. Dies wird als Motivation für die detaillierte und ganzheitliche Betrachtung des Einsatzes von ML-Methoden im Rahmen des IM durch den vorliegenden Beitrag genutzt. Hierbei soll mithilfe einer umfangreichen Literaturanalyse die folgende Forschungsfrage beantwortet werden:

Was ist der State-of-the-Art beim Einsatz von Machine Learning Methoden im Rahmen des Innovationsmanagements?

Ziel dieser Arbeit ist es, den aktuellen Forschungsstand aufzubereiten und hiermit einen Beitrag zum Überblick des Fachgebiets zu leisten. Dabei wird mithilfe einer Konzeptmatrix der Einsatz verschiedener Verfahren entlang des IM untersucht, wodurch eine Orientierungshilfe für weitergehende Forschungsvorhaben entsteht. Zur Erreichung dieses Ziels ist diese Arbeit wie folgt strukturiert. Im zweiten Abschnitt werden zunächst die theoretischen Grundlagen vermittelt, worauf aufbauend die Methodik zur systematischen Literaturanalyse dargestellt wird. In Abschnitt vier werden die Ergebnisse analysiert und veranschaulicht. Abschließend folgt im fünften Abschnitt ein Fazit der gewonnenen Erkenntnisse sowie eine Ableitung von möglichen zukünftigen Forschungsbedarfen.

2 Grundlagen und Begriffe

Der folgende Abschnitt erläutert die zwei Begriffe *Innovationsmanagement* und *Machine Learning Methoden*, um ein einheitliches Begriffsverständnis sowie den nötigen theoretischen Hintergrund zur Beantwortung der Forschungsfrage sicherzustellen.

2.1 Innovation und Innovationsmanagement

Der Innovationsbegriff wird in der Literatur sowohl für den Entstehungsprozess eines neuen Produktes als auch zeitgleich für das neue bzw. verbesserte Produkt selbst verwendet. Im Kontext dieser Arbeit wird die erstere Bedeutung verwendet, wonach allgemein betrachtet, unter dem Begriff Innovation eine neue Art und Weise verstanden werden kann, wie Tätigkeiten durchgeführt und anschließend vermarktet werden [Po90]. Im Detail bedeutet dies die Schaffung neuartiger, deutlich vom Vorgänger abgrenzbarer, Zweck-Mittel-Kombinationen durch bislang unbekannte Produkte, Dienstleistungen und Prozesse bzw. deren Verknüpfungen [Ha16]. In Abhängigkeit des Betrachtungsgegenstands sowie

der Auswirkung auf den Markt lassen sich dabei unterschiedliche Innovationstypologien differenzieren. In der vorliegenden Arbeit wird hierfür die Untergliederung in technische Innovationen, die sowohl Prozesse als auch Produkte beinhalten, *organisationale* sowie *geschäftsbezogene Innovationen* verwendet [ZW95].

Eine Innovation grenzt sich durch die Kommerzialisierung eines Unternehmens von einer Erfindung ab [Ro88]. Folglich lässt sich Innovation als Kernprozess eines Unternehmens definieren, um innerhalb einer volatilen Umgebung zu überleben [SJ07]. Die betrieblichen Aktivitäten zur Steuerung und Ausgestaltung der Innovationsaktivitäten eines Unternehmens werden unter dem Begriff des IM zusammengefasst. Hierunter fallen sowohl strategische Tätigkeiten zur Definition der Innovationsziele wie auch operative Aufgaben, die unter anderem die Planung, Organisation und Kontrolle des Innovationsprozesses sowie die Schaffung entsprechender Rahmenbedingungen innerhalb des Unternehmens beinhalten [PS96]. Nachfolgend werden Aspekte wie bspw. die Innovationskultur nicht weiter thematisiert und der Schwerpunkt der Analyse auf die Tätigkeiten, die im Zusammenhang mit dem Innovationsprozess stehen, gelegt.

Für den Unternehmenserfolg sowie den Erfolg des IM ist es von besonderer Importanz, eine systematische Durchführung des Innovationsprozesses sicherzustellen [GWR09]. Folglich haben sich in der Vergangenheit diverse Vorgehensmodelle etabliert, die sich aufgrund der Diversität hinsichtlich Zielsetzung und Innovationssubjekt legitimieren lassen [Ev10]. Viele Ansätze eint die grundlegende Aufgliederung des Prozesses in die Bereiche Ideen, Projekte und Produkte nach dem Konzept des Innovationstrichters, wonach aus einer Vielzahl von Ideen am Ende des Prozesses eine Handvoll marktreifer Produkte entsteht [FF11]. Im Wesentlichen wird diese Grundidee in dem Dreiphasenmodell des Innovationsprozesses nach [Th92] abgebildet. In der ersten Phase *Ideengenerierung* bestimmt das Unternehmen sein Suchfeld und beginnt mit der Findung bzw. dem Vorschlag von neuen Ideen. Diese werden im zweiten Schritt *Ideenakzeptierung* hinsichtlich ihrer Sinnhaftigkeit für eine potenzielle Umsetzung hin validiert und anschließend Realisierungspläne entwickelt, über welche am Ende der Phase entschieden wird. Fällt die Ideenbewertung positiv aus, werden diese im dritten Prozessschritt, der *Ideenrealisierung*, tatsächlich implementiert und anschließend am Markt vertrieben sowie der Erfolg gemessen. [Th92]. Dieses Prozessmodell repräsentiert mit seinen drei Phasen die wesentlichen Prozessschritte der in [Ev10] untersuchten zwölf verschiedenen Modelle. Zum selben Ergebnis kommen ebenfalls [FF11].

2.2 Machine Learning Methoden

Die steigende Menge an Daten folgert den Wunsch nach automatisierter Verarbeitung, um neue Erkenntnisse sowie Innovationen zu generieren [MG20, Mu12]. In diesem Zusammenhang hat sich innerhalb der Informatik der Teilbereich der KI herausgebildet, welches versucht, mithilfe von Computern intelligentes Verhalten zu simulieren. Es fehlt bislang allerdings an einer einheitlichen, über mehrere Disziplinen anerkannten Definition, da KI

als Sammelbegriff für Datenanalyseprozesse zum Erhalt oder Anwenden von Wissen verwendet wird [Gr14]. Während viele Publikationen der untersuchten Forschungsfrage diesen Begriff nutzen, wird nachfolgend stattdessen auf den Teilbereich des ML genauer eingegangen. Dieser ermöglicht es, ein System mit menschenähnlicher Intelligenz ohne explizite Programmierung zu erzeugen. Hierdurch erhält das System die Möglichkeit, selbstständig in einer spezifischen Aufgabe aus Erfahrungen zu lernen und sich hierdurch stetig zu verbessern [JM15]. Hierbei kann in Abhängigkeit von Grad des Analysewertes die rein beschreibende (*deskriptive*), erklärende (*diagnostische*), vorausschauende (*prädiktive*) sowie optimierende (*präskriptive*) Analyse differenziert werden [Ba20].

Außerdem werden in Abhängigkeit der Aufgabe, die das System löst, sowie der Vorgehensweise innerhalb des Trainings, die nachfolgend dargestellten Typen unterschieden. Erstens wird der Begriff des *überwachten Lernens (engl. supervised learning)* verwendet. Hierunter wird ein Lernalgorithmus verstanden, welcher durch die Abbildung eines bzw. jedes Eingabewertes x auf eine Ausgabe y eine möglichst akkurate Vorhersage trifft. In Abhängigkeit der Ausgabevariablen $y^{(i)}$ kann die Art des Problems bestimmt werden. Ist diese kategorisch, so handelt es sich um eine Klassifizierung oder eine Mustererkennung bzw. eine Anomalieerkennung. Handelt es sich um ein reellwertiges oder ordinales Skalar, so liegt meist ein Regressionsproblem vor [JM15, MRT18, Mu12].

Zweitens wird der Bereich des *unüberwachten Lernens (engl. unsupervised learning)* differenziert. Im Gegensatz zu überwachtem Lernen sind die Vorhersagewerte sowie die Aufgaben unbekannt und können folglich nicht zum Überwachen des Lernalgorithmus verwendet werden. Ein typischer Anwendungsbereich ist das Clustering und die Assoziationsanalyse, wobei darüber hinaus unüberwachtes Lernen bspw. zur Dimensionsreduktion hochdimensionaler Daten, Erkennung von Graphstrukturen, Konzeptbeschreibungen oder Matrixvervollständigung genutzt werden kann [JM15, MRT18, Mu12].

Drittens wird neben den beiden zuvor genannten Verfahren das *bestärkende Lernen (engl. reinforcement learning)* unterschieden. Hierbei wird ein Ergebnis entweder belohnt oder bestraft, um eine entsprechende Anpassung im Training sowie nachgefolgerter Prognose zu erzielen. Im Gegensatz zum überwachten Lernen wird hierbei nicht explizit die korrekte Ein- bzw. Ausgabe präsentiert [JM15, MRT18, Mu12].

Neben den bereits beschriebenen Ansätzen werden heutzutage weitere Mischformen sowie je nach Publikation zusätzliche Verfahren differenziert [JM15]. Hierzu zählen bspw. das selbstüberwachte sowie Transferlernen. Diese sind allerdings in der Praxis aktuell noch weniger verbreitet und werden nachfolgend nicht genauer erläutert [AbPH21].

3 Methodik zur Literaturanalyse

Zur Beantwortung der vorliegenden Forschungsfrage wird die strukturierte Literaturanalyse nach [Fe06] sowie [WW02] verwendet. Gemäß dem Vorgehensmodell von [Fe06] wurde im ersten Abschnitt die Problemdefinition beschrieben, welche durch die folgenden

Schritte Literatursuche und -auswertung, Analyse und Interpretation sowie Präsentation der identifizierten Literatur vervollständigt wird [Fe06]. Für die Literatursuche werden gemäß den Leitlinien geeignete Auswahlkriterien ausgewählt, welche nachfolgend dargestellt werden. Da diese Arbeit den Fokus auf die Darstellung der aktuellen State-of-the-Art legt, werden lediglich Publikationen aus den Jahren 2018 (inkludiert) bis 2022 (soweit vorhanden) betrachtet und auf einen historischen Rückblick verzichtet. Dies deckt sich ebenfalls mit der Publikation von [Ro18], welche im Jahr 2018 noch eine Forschungslücke in diesem Bereich identifiziert haben. Eine weitere Einschränkung wurde durch die Fokussierung auf deutsche sowie englischsprachige Publikationen sowie deren öffentliche Verfügbarkeit gewählt. Für *SpringerLink* wird die zusätzliche Einschränkung auf Conference Paper und Artikel vorgenommen. Als Schlagwörter werden die Begriffe *innovation management* und *machine learning* verwendet und in Form eines Suchterms in die verwendeten Literaturdatenbanken eingegeben. Diese sind: *AIS Electronic Library*, *IEEE Xplore Digital Library*, *SpringerLink* und *Science Direct.* Hierdurch konnten in Summe 398 Publikationen identifiziert werden. Nach Entfernung der Duplikate, Sichtung der Abstracts sowie einer Vorwärts- und Rückwärtssuche konnte diese Anzahl auf 38 Beiträge, die für diese Arbeit relevant sind, reduziert werden. Diese werden im anschließenden Abschnitt in Form einer Konzeptmatrix dargestellt und analysiert.

4 Forschungsergebnisse

Im nachfolgenden Abschnitt werden die Ergebnisse der Literaturanalyse dargestellt. Hierfür wird nach dem Vorgehen von [WW02] eine Konzeptmatrix erstellt, welche die wesentlichen Konzepte, die in Abschnitt zwei dargestellt sind, aufgreift. Dabei bilden die Phasen des IM-Prozesses sowie die dargestellten ML-Verfahren das Zentrum der Konzeptmatrix, um die Problemstellung dieser Forschungsarbeit zu untersuchen. Ergänzend hierzu werden der Ursprung, die Innovationstypologie, die Datenart sowie der Analysetyp und -aufgabe abgetragen. Wird eine der Ausprägungen in der untersuchten Publikation dargestellt oder lässt sich diese aufgrund der Beschreibung folgern, so wird dies durch ein *X* in der Konzeptmatrix dargestellt. Folgend werden die Besonderheiten erläutert und die Konzeptmatrix in Tabelle eins vollständig dargestellt.

4.1 Allgemein

Innerhalb der untersuchten Stichprobe lassen sich lediglich drei Publikationen einem praktischen Ursprung zuordnen. Diese weisen im Kontext der untersuchten Forschungsfrage allerdings keine Unterschiede zu den anderen betrachteten Publikationen auf, weswegen auf eine Differenzierung im weiteren Verlauf verzichtet wird. Eine weitere Betrachtungsdimension im Zusammenhang mit dem Ursprung stellt die Anwendungsdomäne der Publikation dar. In diesem Punkt ist die Fokussierung auf eine spezielle Domäne in 44,7 % der Stichprobe festzustellen, wobei sich fünf Publikationen auf den Bereich der Produktion sowie zwei weitere im Spezifischen der Automobilbranche zuwenden. Setzt man die Publikationen in den zeitlichen Verlauf, so ist ein kontinuierlicher Anstieg innerhalb des

Betrachtungszeitraums festzustellen. Während im Jahr 2018 vier und 2019 fünf Publikationen veröffentlicht worden sind, sind es im Jahr 2020 bereits 11 und im Jahr 2021 16 Arbeiten. Dies veranschaulicht zum einen die Aktualität dieses Themenfeldes und weist außerdem die steigende Forschungsaktivität nach.

	Ursprung			Typologie			IM-Prozess				ML-Verfahren							Datenart				Analysetyp				Analyseaufgabe									
Literaturverweis	Akademisch	Praktisch	Domäne	technische Innovationen	organisationale Innovationen	geschäftsbezogene Innovationen	Ideengenerierung	Ideenakzeptierung	Ideenrealisierung	Phasenunabhängig	überwachtes Lernen	unüberwachtes Lernen	Bestärkende Lernen	Selbstüberwachtes Lernen	Transferlearning	nicht spezifiziert		Natürliche Sprache	Nummerische Daten	Bilder/ Videos	nicht spezifiziert	deskriptive Analyse	diagnostische Analyse	prädiktive Anlyse	präskriptive Analyse	Clusteranalyse	Klassifikation	Regression	Anomalieerkennung	Konzeptbeschreibung	Assoziationsanalyse	Dimensionsreduktion	Visual Analytics	nicht spezifiziert
[AT18]	X		-	X						X	X	X	X				X				X	X	X		X	X	X		X	X		X	
[Ba21]	X		Produktion	X	X					X						X					X												X
[BAE19]	X		Gesundheitswesen	X			X	X	X			X						X				X	X	X		X		X						
[Bl19]		X	Automobilesektor	X					X							X			X	X				X					X					
[BM21]	X		-	X	X		X	X	X							X					X	X	X	X		X						X		
[Bu21]	X		Produktion			X				X						X			X			X	X	X		X			X						
[Da20]	X		-	X			X	X				X						X				X	X			X	X								
[EGC18]	X		Bildungswesen	X			X									X					X			X		X						X		
[ERH19]		X	-		X					X		X			X			X					X			X				X				
[FNv21]	X		-	X				X			X										X			X										X
[Gi18]	X		-	X			X									X			X	X		X				X			X	X				
[Gu20]	X		-	X						X				X	X			X				X				X								
[Ha21]	X		-	X	X		X				X	X	X					X		X		X	X	X		X			X					
[HK21]	X		Sport	X			X	X			X							X	X			X	X			X	X			X				
[Hu21]	X		-	X	X		X									X		X							X	X								X
[JSK21]	X		-	X	X		X	X	X							X		X				X	X			X								
[Kh20]	X		-	X				X			X	X			X			X					X			X	X							
[Kö19]	X		-	X			X					X						X				X				X		X				X		
[KZE20]	X		Energiesektor	X	X				X		X										X		X			X	X							
[LH21]	X		-	X	X				X		X								X					X		X	X	X						
[LHK21]	X		-	X	X					X		X						X				X				X		X		X				
[Li20]	X		Produktion	X						X						X					X			X										X
[LZZ20]	X		-	X				X			X								X					X										X
[MG20]	X		-	X			X				X	X						X						X		X		X				X		
[MUU21]	X		Heritage	X			X				X							X						X			X	X						
[NB19]	X		-	X			X									X		X				X	X	X								X		
[NBK21]	X		-	X			X					X						X				X	X	X		X		X				X		
[Sh18]		X	Bildungswesen	X			X	X	X		X	X						X	X			X		X		X	X			X	X			
[Sj21]	X		Produktion			X				X						X					X					X				X				
[Ta20]	X		Supply Chain	X				X								X					X			X										
[TMT20]	X		-	X				X			X	X						X					X				X	X	X		X			
[Tr21]	X		Personalwesen	X	X				X							X		X	X			X	X	X		X			X	X				
[Ya19]	X		Automobilesektor	X	X		X				X										X			X	X			X		X				
[YX20]	X		-	X	X					X						X					X	X	X			X							X	
[Zh21]	X		Produktion	X			X				X							X				X				X					X	X		
[ZLT21]	X		Supply Chain	X						X		X						X				X		X			X	X						
[ZYC20]	X		Robotik		X					X		X						X				X				X						X		
[ZZB18]	X		-	X			X					X						X	X							X		X				X		

Tab. 1: Konzeptmatrix (eigene Darstellung)

4.2 Forschungsschwerpunkte im IM

Zur Betrachtung des aktuellen State-of-the-Art ist es wichtig, die Publikationen hinsichtlich ihrer Fokussierung innerhalb des IM zu untersuchen. Betrachtet man die Innovationstypologie, so sind fast alle Publikationen (89,5 %) thematisch der technischen Innovation zuordenbar. Einige Publikationen wie bspw. [KZE20, ZYC20] verweisen dabei explizit auf ihre Fokussierung in diesem Bereich, die meisten Publikationen sind dieser Typologie aber durch die thematisierte Prozessoptimierung zuzuordnen. So verweisen zahlreiche Publikationen auf die Potenziale von ML zur Automatisierung der Trendforschung. Hierbei ist es wichtig anzumerken, dass ML bislang nicht für die eigentliche Generierung innovativer neuer Ideen, sondern vielmehr für die Identifikation spannender Ideen aus bisherigen Konzepten eingesetzt wird.

Ein Anteil von 13 Publikationen lässt sich der organisationalen Innovation zuordnen. Hierbei werden übergreifende Themen beschrieben, welche das gesamte Unternehmen prägen. Als Beispiel können [ERH19] gelten, welche ein Bewertungsschema zur Messung des Innovationsfortschritts sowie zur Ableitung weiterer Maßnahmen beschreiben. Ferner betrachten [ZYC20] die Auswahl sowie Suche geeigneter Talente für die Umsetzung von technologischen Innovationsprojekten. 11 (84,6 %) der 13 Publikationen, die dieser Typologie zuzuordnen sind, sind gleichzeitig der technischen Innovation angehörig. Den geschäftsbezogenen Innovationen lassen sich lediglich zwei Publikationen zuordnen. Beide erwähnen keine ML-Verfahren explizit, sondern weisen allgemein auf die Potenziale von KI im Kontext des Business Model Innovation hin. Hierbei verweisen bspw. [Bu21] auf die Analyse von Nutzungsdaten zur Antizipation von Nutzerwünschen und der Ableitung von neuen Funktionen oder Produkten. Gleichzeitig kann potenziell ebenfalls eine Integration von ML-Verfahren im eigentlichen Geschäftsmodell erfolgen, um Konzepte wie dynamische Preisgestaltung oder Massenindividualisierung zu ermöglichen.

Neben der Innovationstypologie lässt sich die Zuordnung zu einem oder multiplen Prozessschritten im IM-Prozess nach [Th92] zur Klassifizierung der Publikationen betrachten. Hierbei muss jedoch berücksichtigt werden, dass 28,9 % der Publikationen keiner Phase zugeordnet werden können. Grund hierfür ist die übergeordnete oder allgemeine Betrachtung des Themenfelds. Die erste Phase der Ideengenerierung wird mit einem besonders hohen Anteil von 47,4 % angewendet. Hierbei wird in den Publikationen vor allem die automatisierte Erkennung von Trends thematisiert. Die folgende Phase der Ideenakzeptierung wird in 28,9 % der Stichprobe thematisiert. Ein Kernthema, das hierbei durch mehrere Publikationen diskutiert wird, ist die Prädiktion des Erfolgs einer Innovation. Sechs Publikationen lassen sich mehr als einer Phase zuordnen. Weiterhin werden hierbei immer mindestens die ersten zwei Phasen zusammen genannt, in drei Fällen sogar alle drei Phasen. Diese betrachten die Anwendung von ML ganzheitlich im Kontext des IM. Exemplarisch können hierfür [JSK21] genannt werden, welche das Einsatzpotenzial von Chatbots entlang des IM-Prozesses ermitteln. Die Ideenrealisierung ist mit einem Anteil von 21,1 % am wenigsten in der Stichprobe vertreten. In diesem Anteil wird mit 62,5 % deutlich häufiger als in der Gesamtstichprobe auf eine spezifische Domäne verwiesen.

4.3 Anwendung von ML-Methoden im Kontext des IM

Zur ganzheitlichen Betrachtung der Forschungsfrage ist es wichtig, neben dem IM ebenfalls die Verwendung von ML in diesem Kontext detailliert zu betrachten. Eine wesentliche Differenzierung kann hierbei durch die Untersuchung des Lernverfahrens erfolgen. Hierbei ist festzustellen, dass im Wesentlichen nur auf die zwei populären Ansätze des überwachten (36,8 %) bzw. unüberwachten Lernens (39,5 %) verwiesen wird. Diese sind nahezu gleich häufig vertreten, wobei keine signifikante Häufung eines Ansatzes für spezifische IM Teilbereiche vorhanden ist. Sechsmal wird gleichzeitig auf beide Verfahren verwiesen. Dies liegt zum einen an den unterschiedlichen Aufgaben, die mithilfe der Verfahren umgesetzt werden und zum anderen an Publikationen wie [Ha21], die einen allgemeinen Überblick über Implementierung liefern. Alle weiteren untersuchten ML-

Verfahren werden zusammengerechnet, lediglich in fünf Publikationen erwähnt und sind folglich für eine weitere Betrachtung nicht relevant.

Eine weitere Informationsquelle ist die verwendete Datenart, da hiermit weitere Informationen zu den durchgeführten Analysen gewonnen werden können. In diesem Zusammenhang ist die hohe Anzahl (60,5 %) an Publikationen auffällig, die auf natürlich sprachliche Texte als Analyseobjekt zurückgreift. Bei genauerer Betrachtung lassen sich diese nochmals in Patente, wissenschaftliche Publikationen sowie weitere Dokumente wie bspw. unternehmensinterne Dokumentationen untergliedern. In der Phase der Ideengenerierung wird diese Datenform insbesondere bei der Trenderkennung zu 77,8 % genutzt. Speziell die Untersuchung von Patenten scheint in diesem Zusammenhang den aktuellen State-of-the-Art abzubilden, da in der Stichprobe neun Publikationen auf diese zurückgreifen. In diesem Punkt sei exemplarisch auf [AT18] verwiesen, welche diesbezüglich eine spezifische Literaturanalyse durchgeführt haben. Numerische Daten als weitere Datenart wurden im Vergleich insgesamt neunmal und Bild bzw. Videodaten dreimal verwendet.

Mit Ausnahme der präskriptiven Analyse, welche lediglich zwei Mal genannt wird, werden die weiteren Analysetypen in einem ähnlichen Umfang verwendet. Innerhalb der untersuchten Publikationen zeigt sich, dass die deskriptive sowie prädiktive Analyse in 52,6 % und die diagnostische Analyse in 42,1 % der Publikationen angewandt wurde. Acht Publikationen nutzen alle drei zuvor genannten Analysetypen. Weiterhin ist bei der Untersuchung der Analyseaufgabe ein besonders hoher Anteil (26 Nennungen) der Clusteranalyse zu erkennen. Gefolgt wird diese von der Regression (12 Nennungen) sowie der Klassifikation (elf Nennungen). Mit acht Nennungen wird die visuelle Analyse sowie die Konzeptbeschreibung gleich oft erwähnt. Die weiteren Aufgaben sowie der Anteil an nicht zuordenbaren Publikationen beträgt fünf.

5 Diskussion und Ausblick

Durch die Analyse der Stichprobe konnten aktuelle Trends im Einsatz von ML für das IM identifiziert werden. Im Allgemeinen ist eine steigende Anzahl an Publikationen zu verzeichnen, wobei diese zumeist die technische Innovation fokussieren. Dies kann darin begründet sein, dass ML häufig in Kombination mit der Optimierung bzw. Automatisierung von Prozessen als Teilgebiet der technischen Innovation in Kontext gebracht wird. Hierbei werden im aktuellen Status-Quo insbesondere die initialen Prozesse des IM betrachtet und dabei die erste Phase der Ideengenerierung unterstützt. Diese Erkenntnis manifestiert sich, da die Publikation von [Ha21] ebenfalls diese Phase in ihrer Analyse fokussieren. Das Handlungsobjekt in diesen Prozessen stellen zumeist Texte in natürlicher Sprache, überwiegend in Form von Patenten, aber auch wissenschaftlichen Publikationen oder sonstigen, internen Unternehmensdaten dar. Schwerpunkte der Analysen sind dabei zumeist die Selektion, Aggregation oder Identifikation interessanter Konzepte innerhalb der Vielzahl an bereits existierenden Dokumenten. Auch im Zusammenhang mit numerischen Daten beschränkt sich der Einsatz von ML auf die Identifikation spannender Zusammenhänge,

die wiederum als Grundlage für die Konzeptionierung neuer Innovationen dienen können. Damit lässt sich in beiden Fällen folgern, dass im aktuellen Status Quo ML den Menschen nur durch Impulse unterstützt, selbstständig allerdings keine Erfindungen erzeugen kann. Stützen lässt sich diese Aussage außerdem durch die Eigenschaften der meistbetrachteten Dokumente, da Patente und wissenschaftliche Publikationen natürlicherweise neuartige Innovationen und Fortschritte verkörpern.

Eine weitere Dimension, die diese Erkenntnis ebenfalls unterstützt, ist die Analyseaufgabe, bei der eine besonders starke Ausprägung auf Clustering-Algorithmen, zur Identifikation neuartiger Trends, festzustellen ist. Hinsichtlich der verwendeten ML-Verfahren ist eine Fokussierung auf die beiden traditionellen Lernverfahren des unüberwachten bzw. überwachten Lernens zu erkennen. In der untersuchten Stichprobe sind die Verfahren des bestärkenden, selbstüberwachten oder Transferlernens bislang noch schwach ausgeprägt. Dies könnte durch die relative Neuheit dieser Ansätze sowie der limitierten Auswahl etablierter Modelle im Vergleich zum überwachten und unüberwachten Lernen erklärt werden. Die Publikation von [HiKa21] verweist in ihrer Future Work auf eine potenzielle Anwendung von BERT, einem neuronalen Netz zur Verarbeitung natürlicher Sprache, das auf dem Ansatz des selbstüberwachten Lernens sowie dem Transferlernen beruht. Folglich ist in Zukunft ein steigender Trend dieser neueren Ansätze zu erwarten.

Zusammenfassend lässt sich die Forschung des aktuellen State-of-the-Art in Bezug auf den Einsatz von ML im IM als gering fortgeschritten bezeichnen. Grund für diese Aussage ist zum einen die recht homogene Fokussierung im Bereich der textbasierten ML-Analyse zur Identifikation relevanter Trends innerhalb existierender Dokumente und zum anderen die geringe bis fehlende Betrachtung weiterer Bereiche des IM innerhalb der untersuchten Stichprobe. Außerdem werden in einigen Publikationen die Potenziale oder die Sinnhaftigkeit von ML-Verfahren in diesem Kontext angesprochen, allerdings nicht weiter konkretisiert oder gar angewendet.

Limitationen der dargestellten Erkenntnisse ergeben sich zum einen aus der beschränkten Stichprobengröße sowie zum anderen aus der uneinheitlichen sowie nicht trennscharfen Definition des Begriffs ML, welcher in den untersuchten Publikationen zumeist ebenfalls als Synonym für KI verwendet wird. Folglich kann eine vollständige Identifikation aller relevanter Publikationen nicht gewährleistet werden. Außerdem bietet die Analyse der Publikationen zur Befüllung der Konzeptmatrix Interpretationsspielraum, wobei die Zuordnung zu den Konzepten nach bestem Wissen und Gewissen erfolgte.

Ein weiterer Forschungsbedarf besteht für die präskriptive Analyse sowie die geschäftsbezogenen Innovationen, da diese bislang beide in lediglich zwei Publikationen thematisiert werden. Da insbesondere bei den geschäftsbezogenen Innovationen bislang keine konkreten ML-Verfahren benannt werden, könnten weitergehende Arbeiten explizit Nutzungsszenarien für die jeweiligen Analyseverfahren beschreiben. Erweitern lässt sich dies um die Betrachtung von bestärkendem, selbstüberwachten oder Transferlernens im Kontext aller Innovationstypologien. Außerdem könnten die genauen ML-Verfahren untersucht werden und eine Validierung dieser Arbeit erfolgen.

6 Literaturverzeichnis

[AbPH21] Abdel-Karim, Benjamin M.; Pfeuffer, Nicolas; Hinz, Oliver: Machine learning in information systems - a bibliographic review and open research issues. In: *Electronic Markets* vol. 31, Electronic Markets (2021), Nr. 3, pp. 643–670

AT18 Aristodemou, L.; Tietze, F.: The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data. World Patent Information October/55, S. 37–51, 2018.

Ba20 Balali, F. et al.: Data Intensive Industrial Asset Management. Data Intensive Industrial Asset Management, S. 105–113, 2020.

Ba21 Babu, M. M. et al.: Exploring big data-driven innovation in the manufacturing sector: evidence from UK firms. Annals of Operations Research 0123456789, 2021.

BAE19 Baskici, C.; Atan, S.; Ercil, Y.: Forecasting of innovation in the light of semantic networks. Procedia Computer Science 158, S. 443–449, 2019.

Bl19 Bloomfield, R. et al.: Disruptive Innovations and Disruptive Assurance: Assuring Machine Learning and Autonomy. Computer 9/52, S. 82–89, 2019.

BM21 Brandtner, P.; Mates, M.: Artificial Intelligence in Strategic Foresight – Current Practices and Future Application Potentials: The 2021 12th International Conference on E-Business, Management and Economics (ICEME 2021). Association for Computing Machinery, New York, NY, USA, S. 75–81, 2021.

Bo21 Boston Consulting Group: Overcoming the Innovation Readiness Gap, 2021.

Bu21 Burström, T. et al.: AI-enabled business-model innovation and transformation in industrial ecosystems: A framework, model and outline for further research. Journal of Business Research June 2020/127, S. 85–95, 2021.

Da20 Daxenberger, J. et al.: ArgumenText: Argument Classification and Clustering in a Generalized Search Scenario. Datenbank-Spektrum 2/20, S. 115–121, 2020.

EGC18 Escandón-Quintanilla, M. L.; Gardoni, M.; Cohendet, P.: Improving concept development with data exploration in the context of an innovation and technological design course. International Journal on Interactive Design and Manufacturing 1/12, S. 161–172, 2018.

ERH19 Eljasik-Swoboda, T.; Rathgeber, C.; Hasenauer, R.: Artificial intelligence for innovation readiness assessment. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 14–19, 2019.

Ev10 Eveleens, C.: Innovation management; a literature review of innovation process models and their implications. Science 2010/800, S. 900–916, 2010.

Fe06 Fettke, P.: State-of-the-art des state-of-the-art: Eine untersuchung der Forschungsmethode "review" innerhalb der Wirtschaftsinformatik. Wirtschaftsinformatik 48/4, S. 257–266, 2006.

FF11 Franken, R.; Franken, S.: Innovationsmanagement. Integriertes Wissens- und Innovationsmanagement, S. 192–298, 2011.

FNv21 Falco, B. D.; Neubert, M.; van der Krogt, A.: Usage of Machine Learning To Predict Market Attractiveness in the Context of Internationalization. Proceedings of the 8Th International Conference Innovation Management, Entrepreneurship and Sustainability (Imes 2020) Imes, S. 140–152, 2021.

Gi18 Giaccardi, E.: Things making things: Designing the internet of reinvented things. IEEE Pervasive Computing 3/17, S. 70–72, 2018.

Gr14 Grewal, D. S.: A Critical Conceptual Analysis of Definitions of Artificial Intelligence as Applicable to Computer Engineering. IOSR Journal of Computer Engineering 2/16, S. 9–13, 2014.

Gu20 Guarino, G. et al.: SummaTRIZ Summarization Networks for Mining Patent Contradiction. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 979–986, 2020.

GWR09 Gaubinger, K.; Werani, T.; Rabl, M.: Praxisorientiertes Innovations- und Produktmanagement. Praxisorientiertes Innovations- und Produktmanagement, S. 4–16, 2009.

Ha16 Hauschildt, J. et al.: Innovationsmanagement. Vahlen, 2016.

Ha18 Hanschke, I.: Digitalisierung und Industrie 4.0-einfach und effektiv: Systematisch und lean die Digitale Transformation meistern. Carl Hanser Verlag GmbH Co KG, München, 2018.

Ha21 Haefner, N. et al.: Artificial intelligence and innovation management: A review, framework, and research agenda. Technological Forecasting and Social Change October 2020/162, S. 120392, 2021.

HK21 Hippel, E. von; Kaulartz, S.: Next-generation consumer innovation search: Identifying early-stage need-solution pairs on the web. Research Policy 8/50, S. 104056, 2021.

Hu21 Hutchinson, P.: Reinventing Innovation Management: The Impact of Self-Innovating Artificial Intelligence. IEEE Transactions on Engineering Management 2/68, S. 628–639, 2021.

JM15 Jordan, M. I.; Mitchell, T. M.: Machine learning: Trends, perspectives, and prospects. Science 6245/349, S. 255–260, 2015.

JSK21 Johannsen, F.; Schaller, D.; Klus, M. F.: Value propositions of chatbots to support innovation management processes. Information Systems and e-Business Management 1/19, 2021.

Kh20 Khiat, A.: Learning a lightweight representation: First step towards automatic detection of multidimensional relationships between ideas. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 230–233, 2020.

KZE20 Khalyasmaa, A. I.; Zinovieva, E. L.; Eroshenko, S. A.: Intelligent Data Analytics System for Innovation Efficiency Assessment Based on Risk Analysis. 2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT), S. 479–482, 2020.

LH21 Luo, X.; Huang, R.: Application of Adaptive Co-evolutionary Algorithms to Technology Innovation Management. Wireless Personal Communications 0123456789, 2021.

LHK21 Lee, C; Hong, S.; Kim, J.: Anticipating multi-technology convergence: a machine learning approach using patent information. Scientometrics 3/126, S. 1867-1896, 2021.

Li20 Liu, J. et al.: Influence of artificial intelligence on technological innovation: Evidence from the panel data of china's manufacturing sectors. Technological Forecasting and Social Change May 2019/158, 2020.

LZZ20 Li, H.; Zhang, Q.; Zheng, Z.: Research on enterprise radical innovation based on machine learning in big data background. Journal of Supercomputing 5/76, S. 3283–3297, 2020.

MG20 Mühlroth, C.; Grottke, M.: Artificial Intelligence in Innovation: How to Spot Emerging Trends and Technologies. IEEE Transactions on Engineering Management 2/69, S. 493–510, 2020.

MRT18 Mohri, M.; Rostamizadeh, A.; Talwalkar, A.: Foundations of machine learning. MIT press, 2018.

Mu12 Murphy, K. P.: Machine learning: a probabilistic perspective. Chance Encounters: Probability in Education, 2012.

Mü20 Mühlroth, C.: Artificial Intelligence as Innovation Accelerator: Proceedings of the 2020 on Computers and People Research Conference. Association for Computing Machinery, New York, NY, USA, S. 6–7, 2020.

MUU21 Münster, S.; Utescher, R.; Ulutas Aydogan, S.: Digital topics on cultural heritage investigated: how can data-driven and data-guided methods support to identify current topics and trends in digital heritage? Built Heritage 1/5, 2021.

NB19 Nazemi, K.; Burkhardt, D.: A Visual Analytics Approach for Analyzing Technological Trends in Technology and Innovation Management. In (Bebis George et al. (eds) Hrsg.): Advances in Visual Computing. ISVC 2019. Lecture Notes in Computer Science, vol 11845. Springer, Cham. Springer International Publishing, S. 283–294, 2019.

NBK21 Nazemi, K.; Burkhardt, D.; Kock, A.: Visual analytics for technology and innovation management. Multimedia Tools and Applications, 2021.

Po90 Porter, M.: Competitive Advantage of Nations. Competitive Intelligence Review 1/1, S. 14, 1990.

PS96 Pleschak, F.; Sabisch, H.: Innovationsmanagement. Stuttgart Schäffer-Poeschel, 1996.

Ro18 Rossetto, D. E. et al.: Structure and evolution of innovation research in the last 60 years: review and future trends in the field of business through the citations and co-citations analysis. Scientometrics 3/115, S. 1329–1363, 2018.

Ro88 Roberts, E. B.: What We've Learned: Managing Invention and Innovation. Research-Technology Management 1/31, S. 11–29, 1988.

SB22 Szafarski, D.; Beckmann, H.: Einsatz von Machine Learning im Innovationsmanagement – Eine Analyse des State-of-the-Art. AKWI 2022 –Tagungsband der 35. AKWI Jahrestagung 2022, S. 1-13, 2022.

Sh18 Shams, R.: Developing Machine Learning Products Better and Faster at Startups. IEEE Engineering Management Review 3/46, S. 36–39, 2018.

SJ07 Stern, T.; Jaberg, H.: Erfolgreiches Innovationsmanagement: Erfolgsfaktoren - Grundmuster - Fallbeispiele, 2007.

Sj21 Sjödin, D. et al.: How AI capabilities enable business model innovation: Scaling AI through co-evolutionary processes and feedback loops. Journal of Business Research April/134, S. 574–587, 2021.

Ta20 Tarasova, V. N. et al.: Model of intelligent system for quality evaluation of services on design stage (transport and logistic business as an example). 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 294–298, 2020.

Th92 Thom, N.: Innovationsmanagement. Schweizerische Volksbank, 1992.

TMT20 Tan, J.; Miao, D.; Tan, Q.: Empirical Study on Influencing Factors of Knowledge Product Remixing in OIC. IEEE Access 8, S. 34215–34224, 2020.

Tr21 Trocin, C. et al.: How Artificial Intelligence affords digital innovation: A cross-case analysis of Scandinavian companies. Technological Forecasting and Social Change July/173, S. 121081, 2021.

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.

Ya19 Yamamura, C. L. et al.: The front-end of product development as systems thinking and predictive learning. Procedia Manufacturing 2019/39, S. 1346–1353, 2019.

YX20 Yu, S.; Xiao, Q.: Research on the Construction of Strategic Decision-making Platform of Science and Technology Think Tank Based on Multi-agent Collaboration. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 319–324, 2020.

Zh21 Zhou, Y. et al.: A deep learning framework to early identify emerging technologies in large-scale outlier patents: an empirical study of CNC machine tool. Wireless Personal Communications 2/126, S. 969–994, 2021.

ZLT21 Zhang, Q.; Liu, S.; Tu, Q.: Knowledge organization of node enterprises' technological innovation under supply chain environment. Complex & Intelligent Systems 0123456789, 2021.

ZW95 Zahn, E.; Weidler, A.: Integriertes Innovationsmanagement. Handbuch Technologiemanagement, Stuttgart 1995, S. 351, 1995.

ZYC20 Zhao, N.; Yang, G.; Cao, Y.: Mining Technological Innovation Talents Based on Patent Index using t-SNE Algorithms*: Take the Field of Intelligent Robot as an Example. 2019 IEEE International Symposium on Innovation and Entrepreneurship, TEMS-ISIE 2019, S. 595–601, 2020.

ZZB18 Zhao, Y.; Zhou, C.; Bellonio, J. K.: New value metrics using unsupervised machine learning, lexical link analysis and game theory for discovering innovation from big data and crowd-sourcing. IC3K 2018 - Proceedings of the 10th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management 2, S. 327–334, 2018.

Konzeption eines Machine-Learnings-Verfahrens zum Lösen von Green Vehicle Routing Problemen

Pablo Stockhausen [1], Andreas Johannsen[2], Robert Maurer[3]

Abstract: Dieser Beitrag entwickelt ein Konzept zur praktischen Umsetzung eines Machine-Learning-Verfahrens zum Lösen von Vehicle Routing Problemen im Kontext einer nachhaltigen "Letzte-Meile"-Logistik, welches durch einen Prototyp umgesetzt und getestet wurde. Dabei wurden Aspekte von kombinatorischen Optimierungsalgorithmen in Form eines Ameisenalgorithmus zur Unterstützung des angewendeten Machine-Learning-Systems verwendet. Der Prototyp basiert auf einem "Reinforcement Learning"-System und verwendet als Algorithmus "REINFORCE mit Baseline". In einer Vergleichsanalyse wird der Prototyp mit dem bekannten Vertreter für kombinatorische Optimierungsalgorithmen, Google-OR, an Hand von zwei Anwendungsszenarien verglichen. Die kombinatorischen Optimierungsalgorithmen konnten sich hinsichtlich der Lösungsqualität gegenüber dem Prototyp durchsetzen. Dafür überzeugt der Prototyp in der Laufzeit und dem Automatismus. Es wurde festgestellt, dass der verwendete Machine-Learning-Kontext für kleine bis mittelgroße Logistikdomänen nur geringe Vorteile ergibt. Eine Verwendung von lernenden Systemen für das Vehicle Routing Problem empfiehlt sich bei einem größeren Stoppvolumen und einer erweiterten IT-Infrastruktur. Letztlich bietet der Prototyp in diesem Beitrag eine Alternative gegenüber dem Outsourcing an Drittanbietern zum Lösen von Routingproblemen.

Keywords: Vehicle Routing Problem; Machine Learning; Reinforcement Learning; Ant Colony Optimization; Letzte- Meile- Logistik; Ameisenalgortyhmus.

1 Vehicle Routing Probleme im Kontext der „Letzten-Meile"-Logistik

1.1 Motivation und Zielsetzung

Der vorliegende Beitrag wurde in erster Version auf der AKWI 2022 vorgestellt [SJM22]. Der vorliegende Beitrag wurde in erster Version auf der AKWI 2022 vorgestellt [SJM22].Im Jahr 2020 umfasste das Sendungsvolumen in Deutschland 4,05 Mrd. Paket-, Express- und Kuriersendungen mit einer in den kommenden Jahren steigenden Tendenz, wobei der Hauptanteil an Sendungen den B2C Bereich bedient hat [BU21]. Zu dem Anstieg an Paketsendungen und den damit auftretenden

[1] Universität Potsdam, Wirtschafts- und Sozialwissenschaftlichen Fakultät, August-Bebel-Straße 89, 14482 Potsdam, stockhausen017@gmail.com
[2] Technische Hochschule Brandenburg, Fachbereich Wirtschaft, Magdeburger Str. 50, 14770 Brandenburg an der Havel, johannse@th-brandenburg.de>
[3] Technische Hochschule Brandenburg, Fachbereich Wirtschaft, Magdeburger Str. 50, 14770 Brandenburg an der Havel, robert.maurer@th-brandenburg.de

doi:10.18420/inf2022_75

Routingproblemen, ist auch das wissenschaftliche Interesse für Vehicle Routing Probleme (VRP) auf Basis von modernen IT-Lösungen gestiegen. Eine Suche nach den Begriffen „Vehicle Routing“ in der Datenbank IEEE Explore (https://ieeexplore.ieee.org“) ergab über 6500 Treffer in den letzten zehn Jahren. Ungeachtet dessen, gibt es nur wenige wissenschaftliche Arbeiten im Bereich der Machine-Learning (ML) -Verfahren zum Lösen von VRP. Eine Suche mit Google Scholar (https://scholar.google.com) ergab 15 Veröffentlichungen in den letzten fünf Jahren, welche im Titel die Begriffe „Vehicle Routing“ und „Machine-Learning“ verwendet haben. Zusätzlich bewegen sich die fundamentierten Forschungserkenntnisse in einem überwiegenden theoretischen Umfeld mit Schwerpunkt auf mathematischen Modellen und Annahmen, jedoch mit wenig Praxisbezug. Die Hauptmotivation dieser Arbeit liegt darin, ein praktisches Konzept, in Form eines Prototyps, zum Lösen von VRP zu entwickeln und im Besonderen die Aspekte eines Capacitated Vehicle Routing Problems (CVRP) zu berücksichtigen. Dieser Beitrag verwendet CVRP und VRP gleichbedeutend. Weiterhin soll gezeigt werden, dass sich ML-Verfahren in einem produktiven Umfeld durchaus beweisen und eine langfristige Alternative zum Outsourcing an Drittanbieter darstellen können. Dies wird durch einen Vergleich, hinsichtlich der Effektivität unter ausgewählten Bewertungskriterien, zwischen ML-Verfahren und den klassischen Optimierungsalgorithmen zum Lösen von VRP dargestellt.

1.2 Thematische Hinführung

Das Vehicle Routing Problem (VRP) beschreibt im Gebiet der emissionsfreien „Letzte-Meile“-Großstadtlogistik ein kombinatorisches Optimierungsproblem, welches folgende Grundfrage behandelt: „Was ist der optimale Satz von Routen für einen bestimmten Fuhrpark, um eine bestimmte Anzahl von Kunden zu beliefern?“ [BRN15, S.1]. Das VRP wurde erstmalig von Dantzig & Ramser (1959) durch ihre wissenschaftliche Arbeit „The Truck Dispatching Problem“ diskutiert, wobei der Problemkontext die Lieferung von Kraftstoffen darstellte und mithilfe von algorithmischen Wegen gelöst wurde [DR59]. Zum Lösen von VRP eignen sich besonders Machine-Learning (ML)-Verfahren, da sie algorithmisch vorgehen und erfahrungsbasiert entscheiden können. Die Erfahrungen ergeben sich durch die in bestimmten Themendomänen auftretenden Abonnement-Strukturen und den daraus resultierenden bekannten VRP-Instanzen. Der spezifische Anwendungsfall des VRP entsteht im Rahmen dieser Arbeit durch den Kontext einer „Letzten-Meile“-Großstadtlogistik. Dies ist eine moderne Form der urbanen Logistik und umfasst den letzten Logistik Schritt einer Paketzustellung. Im Besonderen wird ein emissionsfreier Ansatz angestrebt, der durch die Verwendung von Lastenrädern sowie Mikro-Depots realisiert wird und die Bezeichnung als Green-VRP begründet [RSC20, S. 6]. Der Aspekt einer „Letzten-Meile“-Großstadtlogistik spielt beim Lösen von kombinatorischen Optimierungsproblemen dahingehend eine Rolle, dass in der Regel von kurzen Routendistanzen ausgegangen werden kann und Touren durch die Verwendung von Lastenfahrrädern klaren Kapazitätsbeschränkungen

unterliegen. Die Planung effizienter Touren ist somit nicht nur ein ausschlaggebender Faktor für das Ausmaß der hohen Logistikkosten einer emissionsfreien Lieferung, sondern auch nötig, um mit der Alternative von Diesel- bzw. Benzin-Kleintransportern mithalten zu können.

1.3 Herausforderungen & Wettbewerbsfähigkeit

Es bestehen Herausforderungen in der späteren Implementierung des Prototyps bezüglich des Wettbewerbs zwischen klassischen Optimierungsalgorithmen und des verwendeten ML-Verfahrens zum Lösen von VRP. Besonders im Bezugspunkt der Performance wird das gewählte ML-Verfahren des Prototyps anfänglich fordernder sein, da es mehr Prozessschritte, wie beispielsweise das Trainieren des Modells, durchlaufen muss. Der Prototyp ist dann wettbewerbsfähig, wenn: Er kurze Routen auf Basis der Tour länge bildet, die Ergebnisse in wenigen Sekunden bereitstellt, allgemein auf nicht bekannte VRP-Instanzen reagiert, auf allen VRP-Instanzen gleich gut arbeitet und kein manuelles Eingreifen benötigt wird.

2 Methodik

2.1 Auswahl des Machine-Learning -Verfahrens

Im Rahmen dieses Beitrages verwendet das ML-Modell ein Reinforcement Learning (RL)-System zum Lösen von VRP der „Letzte-Meile“-Großstadtlogistik [RA18]. Das RL unterscheidet sich maßgeblich von den Alternativen des Supervised- und Unsupervised-Learnings, da es eine andere Herangehensweise für den Aufbau eines Lernenden-Systems verwendet. Das lernende Modell beschreibt im RL einen Agenten oder auch Entscheidungsträger, der ein Environment beobachtet, Aktion darauf ausführt und deren Dynamiken, beginnend von einem unwissenden Startpunkt eigenständig lernt [RA18, S. 2]. Grundlegend steht $A_{(st)} \subseteq A$ für die Menge der Aktionen, die der Agent zu einem Zeitpunkt t ausführen kann [Lo20, S.15]. Dabei beschreibt $a_t \subseteq A$ die ausgewählte Aktion in einem Zeitpunkt t des Agenten. Nach jeder Aktion kann der Agent zwei Arten von Belohnungen erhalten: Eine sofortige oder eine verzögerte Belohnung. Eine sofortige Belohnung findet Anwendung bei Aktionen des Agenten, die eine unmittelbare Beurteilung zu lassen. Diese wurde auch in dem entwickelten Prototyp verwendet. Zum Beispiel lässt sich das Überqueren einer roten Ampel unmittelbar bewerten, da ein negatives Verhalten direkt erkennbar ist und nicht erst durch Folgeverhalten erkenntlich wird. Unter einer verzögerten Belohnung kann beispielsweise eine Aktion in einer Schach- oder Go-Partie verstanden werden, weil die Belohnung sich an den Folgereaktionen misst. Diese wurde auch in der wissenschaftlichen Arbeit von Silver et al. (2017) verwendet [Si17]. Es wurde sich für RL-System entschieden, um ohne

menschliches Vorwissen, Lösungen für komplexe Optimierungsprobleme zu finden und somit Entwicklungs- und Dispositionsaufwand einzusparen.

2.2 Trainings- und Ergebnisdaten

In Kooperation mit einem Unternehmen, welches sich auf nachhaltige Transportlösungen spezialisiert hat, konnten als Datengrundlage die anonymisierten Kundendaten der Ökobox-Anbieter in Berlin zum Trainieren und Testen des Modells verwendet werden. Die Datenobjekte verfügen über folgende Eigenschaften: Längengrad, Breitengrad, Stoppgewicht, Stoppvolumen und einem anonymisierten Identifikator. Die Datenobjekte sind für ein ausgewähltes Datum und einem bestimmten Micro-Depot spezifisch.

2.3 Prototyping

Zur Fundamentierung der Forschung und auf Basis der praktischen Motivation wurde ein qualitativ-konstruktivistischer Ansatz in Form des Prototypings ausgewählt. Der Prototyp soll in der Endfassung einen Service darstellen, der ML-Modelle trainiert, speichert und anwendet. Das ML-Modell ist gegenüber dem Versender, Micro-Depot, Lieferant, Wochentag, Fahrzeuggewicht, Fahrzeugvolumen sowie ML-Verfahren sensibel und wird auf Instanz Basis gespeichert. Das ML-Modell wird nur durch Stopps von einem Tag bis zu einer Woche trainiert, um zu vermeiden, dass Abonnement-Kunden durch doppeltes Auftreten die Strategie des Agenten verfälschen.

2.4 Vergleichsanalyse

Im Kontext dieses Beitrags wird eine Vergleichsanalyse zwischen dem entwickelten Prototyp und dem VRP-Solver von Google-OR durchgeführt. Der VRP-Solver basiert nach Google (2021) auf heuristischen Algorithmen, die als „First Solution Strategy“ kategorisiert werden und optional durch „Local-Search“-Strategien erweitert werden können [Go21]. Die Resultate aus dem Vergleich sollen den Prototypen evaluieren, Verbesserungsmöglichkeiten aufdecken und das Potenzial einer Implementierung von Reinforcement Learning (RL)-Verfahren zum Lösen von VRP in Erwägung ziehen. Die Analyse ist in zwei Teile gegliedert. Im ersten Teil der Vergleichsanalyse liegt der Hauptfokus auf einer Effektivitätseinschätzung des entwickelten Prototyps mit der zugrundeliegenden Basisfunktionalität des ausgewählten RL-Algorithmus „REINFORCE mit Baseline“ zum Lösen von VRP. Hiernach beruht die Effektivitätseinschätzung auf den Schwerpunkten der Trainings- sowie Testlaufzeit, aber auch der Vergleichskriterien der Distanz und des zeitlichen Aufwandes. Der Distanzaufwand wird mit der Haversine-Formel belegt. Hierbei wird der Zeitaufwand durch die hinterlegten Fahrzeuggeschwindigkeiten, die Distanz und einer definierten Stoppverweildauer von fünf Minuten errechnet. Weitere Vergleichskriterien für ein

aufbauendes Experiment wären die Profitabilität, Service-Qualität, Konsistenz sowie externe Faktoren [VLM19, S. 2]. Es wird ein Stoppvolumen von 430 Stopps verteilt auf vier Wochen im Mai 2021 betrachtet, wobei das ML-Modell nach jedem Tag mit 200 Iterationen trainiert wird. Es wird aufbauend trainiert, sodass jede folge Woche auf die Erfahrung der Vorangegangenen aufbaut. Im zweiten Teil der Analyse liegt das Augenmerk auf dem Verbesserungspotenzial des ML-Verfahrens und dem Ausarbeiten von Anpassungsimpulsen, die aus dem vorherigen ersten Teil abgeleitet werden. Das Ziel ist es, sich so konvergent wie möglich an den errechneten Distanz- sowie Zeitaufwand der Konkurrenz anzunähern, um dadurch mit einer besseren Laufzeit zu überzeugen. Im zweiten Teil wird derselbe Problemkontext für einen aussagekräftigeren Vergleich betrachtet. Hierbei wird analog zum ersten Teil nach jedem Tag mit angepassten 115 Iterationen trainiert. Die Iterationsanzahl wurde auf 115 herabgesetzt, um mögliches „overfitting“ des ML-Modells gegenüber den Trainingsdaten zu verhindern.

3 Ergebnisse der Forschungsarbeit

3.1 Modellierung und Konzeption des Machine-Learning-Modells

Das ML-Modell beruht grundlegend auf einem Markov Decision Process (MDP) in Verbindung mit einem Ant Colony Optimization (ACO)-Algorithmus. Dies ist ein Verfahren der kombinatorischen Optimierung, wobei die Grundsteine des Modells durch Hyeong Soo Chang et al. (2004) gelegt wurden [Ch04]. Das ACO-Verfahren wurde im Trainingsmodus des Prototyps als erste Prozessinstanz eingesetzt, um dem RL-Verfahren im anschließenden Prozessschritt eine Vorahnung der Übergangswahrscheinlichkeitsverteilung bereitzustellen und schon trainierte ML-Modelle für den betrachteten Problemkontext einzustimmen. Vergleichsweise wurde ein ähnliches Unterfahren von Yuan Sun et al. (2020) durchgeführt, welche aber als Ausgangslage ein durch ML unterstützendes ACO-Modell betrachtet haben [Su20]. Das ML-Modell wird durch ein MDP definiert. Bei Markov-Modellen sind der Folgezustand und die RL-Belohnung nur vom momentanen Zustand und der gewählten Aktion des Agenten abhängig [Lo20, S.15]. Das angepasste MDP in dem entwickelten Prototyp wurde in Anlehnung an Puterman (2014) modelliert [Pu14]. Als ML-Algorithmus verwendet der Prototyp den „REINFORCE“-Algorithmus oder auch Monte-Carlo Policy-Gradient, um eine optimale Policy π^* zu finden. Für die Entwicklung des Prototyps und der Anwendung des „REINFORCE“-Algorithmus wurde sich an dem Buch „Reinforcement Learning – An Introduction“ von Sutton & Barto (2018) orientiert [SB18]. Dabei zeichnen sich Monte-Carlo-Methoden dadurch aus, dass kein ganzheitliches Wissen über das Environment für die Findung einer optimalen Policy nötig ist. Dies liegt daran, dass das System von den Interaktionen mit dem Environment lernt [SB18, S. 91]. Der „REINFORCE“-Algorithmus des Prototyps baut auf einer parametrisierten Policy $\pi\ (a \mid s, \theta)$ auf und verwendet zusätzlich einen Vergleich

zwischen den momentanen Aktionswerten und einer willkürlichen Baseline. In abgewandelter Form ist der „REINFORCE mit Baseline“-Algorithmus nach Sutton & Barto (2018) konstruiert und wurde in Abb. 1 dargestellt [SB18, S. 330]. Eine Baseline kann dabei eine zufällige Variable sein oder wie im Rahmen dieses Prototyps eine „state-value“-Funktion, die eine Bewertung über die möglichen Belohnungen des gesamtem Zustandsraumes widerspiegelt [SB18, S. 329].

Data : Iterationen k, parametrisierte Policy $\pi(a|s,\theta)$, parametrisierte „state-value“-Funktion (baseline) $V(s,w)$
Result : optimal parametrisierte Policy π^*
initialisieren der Policy-Parameter θ;
initialisieren der Gewichte w;
for $i \leftarrow 0$ **to** k **by** 1 **do**
  $Eps \leftarrow$ Episode nach $\pi(\cdot|\cdot,\theta) = \{S_0, A_0, R_1, ..., S_{T-1}, A_{T-1}, R_T\}$;
  foreach *step* $t \in Eps$, $t = 0, 1, ..., T-1$ **do**
    $G \leftarrow \sum_{j=t+1}^{T} \gamma^{j-t-1} \times R_j$;
    $\delta \leftarrow G - V(S_t, w)$;
    $w \leftarrow w + \alpha \times \gamma^t \times \delta \times \nabla V(S_t, w)$;
    $\theta \leftarrow \theta + \alpha \times \gamma^t \times \delta \times \nabla ln\pi(A_t|S_t, \theta)$;
  end
end

Abb. 1 „REINFORCE mit Baseline“-Algorithmus in Anlehnung an Sutton und Barto (2018) [SB18, S.330]

3.2 Prototypenentwicklung

Das Backend des Prototyps wird in der Programmiersprache Python entwickelt und das Web-Frontend, welches zum besseren Debugging nötig ist, in React. Es wurden bewusst keine vorhandenen ML-Frameworks wie beispielsweise Keras, Tensorflow oder PyTorch verwendet. Die Entscheidung liegt darin begründet, dass im Falle der Verwendung von komplexen ML-Frameworks das Risiko einer Abhängigkeit von dem jeweiligen Framework-Support bestünde. Darüber hinaus zeigt sich ein besserer Verständnis- und Erfahrungsgewinn bei einer Eigenproduktion von komplexen ML-Verfahren gegenüber dem Einsetzen verschiedener ML-Frameworks, wobei eher dazu tendiert wird, das Framework zu lernen und nicht das darauf aufbauende ML-Verfahren. Weiterhin traten bei der Entwicklung des Prototyps spezielle Herausforderungen auf. Anführend dafür ist die Thematik des lokalen Minimums, in dieser sich der Prototyp in den Anfangsphasen der Entwicklung häufig verfangen hat. In Abb. 2 ist ein Beispieldurchlauf der kumulierten Belohnungen, welche als Gesamtlänge in km aller

gebildeten Touren in einer Episode verstanden werden, über 2000 Episoden dargestellt. In diesem Fall betrachtet der Agent ein Problemkontext von 19 Stopps verteilt in Berlin. Dabei findet er zum Ende hin eine gute Lösung, wobei ihm sich die Beste Lösung nicht erschließt, obwohl er diese bereits in den ersten 250 Iterationen entdeckt hatte. Dies lag zum einen daran, dass der Lernfaktor zu niedrig eingestellt war, aber zum anderen der Erkundungsfaktor den Agenten nicht genug zum erweiterten Erkunden des Environments beeinflusst hatte. Zur besseren Erkundung wurde sich einer angepassten dynamischen Epsilon-Erkundung bedient [SB18, S.30].

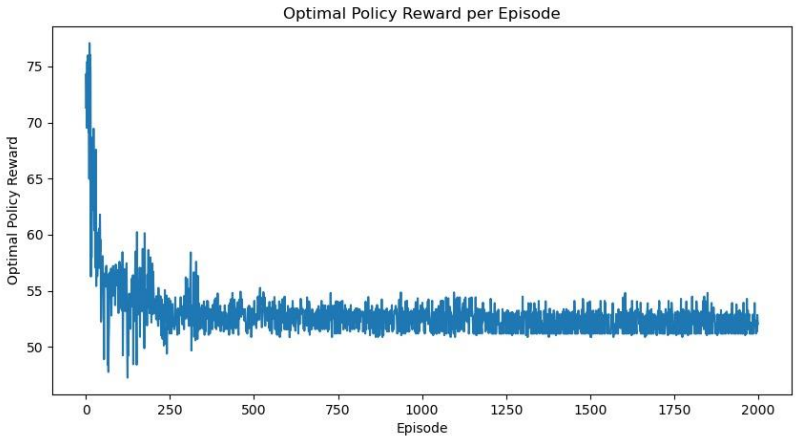


Abb. 2 Beispiel Durchlauf eines ML-Verfahren im lokalen Minimum

4 Interpretation

4.1 Evaluation der Ergebnisse

Im ersten Teil der Vergleichsanalyse und der verwendeten Basis Implementation ist zu erkennen, dass das ML-Verfahren in den ausgewählten Bewertungspunkten schlechter abgeschlossen hat als der Konkurrent. Ein Durchsetzen anhand der Laufzeit ist zwar möglich, dafür sollten aber die Touren nicht deutlich mehr Distanzaufwand in Anspruch nehmen. Die Bezahlung im Bereich der „Letzten-Meile“-Logistik erfolgt in der Regel pro ausgefahrenen Stopp. Aufgrund dessen wollen die Fahrer nicht durch schlecht optimierte Routen gebremst werden. Der Prototyp mit der Grundlagen Implementation von „REINFORCE mit Baseline“ erkennt im ersten Analyseschritt teilweise nahe gelegene Stopps nicht und wählt dadurch überflüssig längere Strecken aus. Er weicht im Durchschnitt +6,94 km und +14,13 min von der Lösung des Konkurrenten ab. Die Resultate des zweiten Analyseschritts zeigen eine Steigerung in den Bewertungspunkten

gegenüber der ersten Version. Die Abweichungen der kumulierten Distanzen für die gebildeten Touren an den jeweiligen Tagen sind in Abb. 3 dargestellt. Die durchschnittlichen Abweichungen zwischen der Prototyp Lösung und OR Strategie liegt bei +6,51 km +11,01 min.

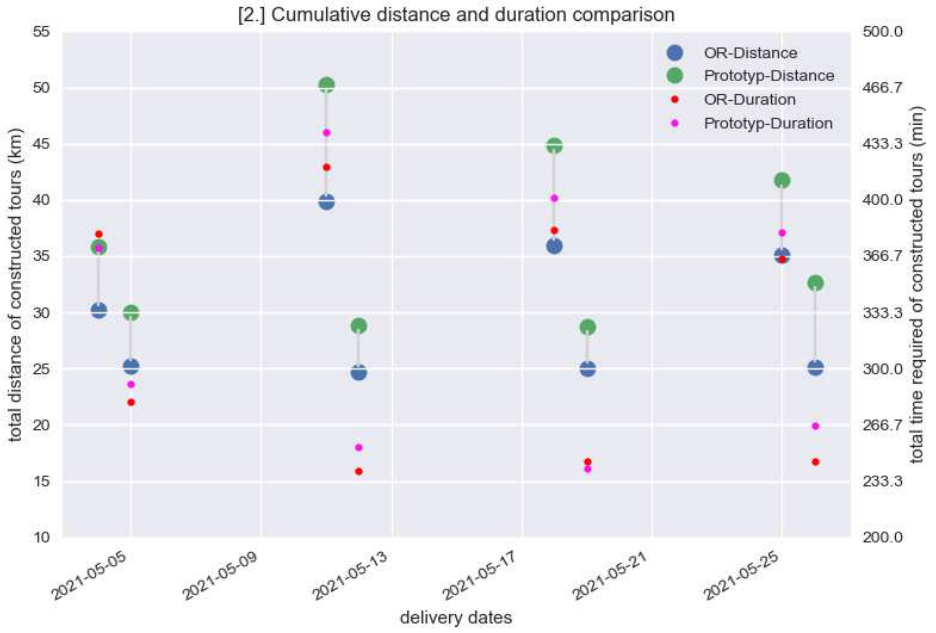


Abb. 3 Vergleich des kumulierten Distanz- und Zeitaufwandes zwischen OR Lösung und erweiterter Prototyp Lösung im zweiten Analyseschritt

Der Prototyp ist auch im zweiten Teil der Analyse bezüglich der Laufzeit zum Lösen von VRP-Instanzen deutlich schneller als der Konkurrent, weil er erfahrungsbasierte Entscheidungen treffen kann. Der nötige zeitliche Trainingsaufwand für den zweiten Analyseschritt wurde in Abb. 4 dargestellt.

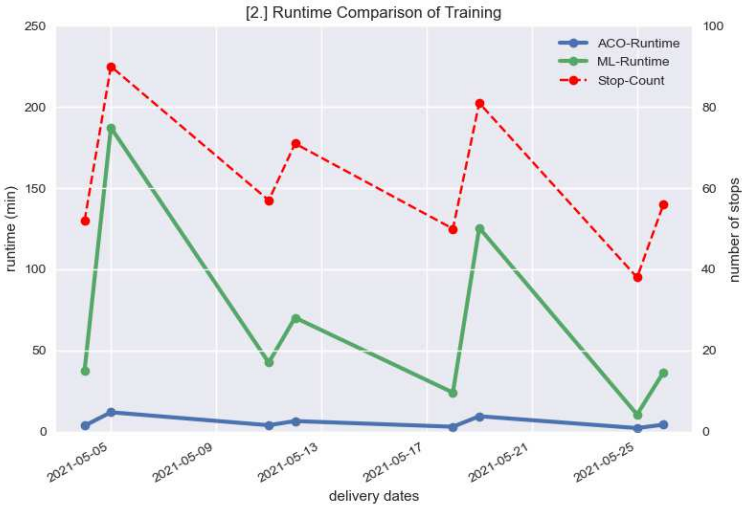


Abb. 4 Vergleich des Trainings-Laufzeitverhaltens zwischen OR und der Prototyp Lösung im zweiten Analyseschritt

Dieser ist für eine produktive Umgebung in Relation zu der Stoppanzahl noch zu aufwendig und sollte im Hinblick auf die Wettbewerbsfähigkeit optimiert werden. In Hinsicht auf die genannten Herausforderungen und damit indirekten Anforderungen an die Wettbewerbsfähigkeit des Prototyps, im Gliederungspunkt 1.3, lassen sich folgende resultierende Punkte, welche der Prototyp in der Lage ist zu erfüllen bzw. nicht zu erfüllen, ableiten:

- Nach Durchlauf des Trainingsprozesses ist der Prototyp in der Lage eine gute bis sehr gute Lösung zu finden.
- Der Prototyp kann ohne manuelles Eingreifen Lösungen erarbeiten.
- Im Testmodus wird das Ergebnis des Prototyps in Millisekunden bereitgestellt.
- Der Prototyp schneidet optimal auf hauptsächlich bekannten VRP-Instanzen ab. Sofern der überwiegende Teil der VRP-Instanz dem Prototyp nicht bekannt ist, wird der Prototyp keine annähernd optimale Lösung finden.
- Aufgrund dessen, dass der Prototyp auf nicht bekannte VRP-Instanzen schlechter reagiert, ist es auch nicht gegeben, dass er auf allen VRP-Instanzen grundlegend gleich arbeitet.

4.2 Vergleich zu kombinatorischen Optimierungsalgorithmen

Die Vergleichsanalyse ergab, dass das ML-Verfahren in seiner jetzigen Implementation nicht das gleiche Niveau der Lösungsqualität für VRP-Instanzen aufweisen kann wie Optimierungsalgorithmen. Des Weiteren zeigten die ausgewählten Anpassungsimpulse für die Basis Implementation von "REINFORCE mit Baseline" keine signifikante Verbesserung. Die Anpassungsimpulse erreichten das der Prototyp in weniger Iterationen eine bessere Lösung erzielte. Die Verbesserung der Lösung weicht aber nur geringfügig von der Basis Implementation ab. Ungeachtet dessen weisen Vergleichsstudien von RL-Verfahren mit verschiedenen Bewertungskriterien und gleicher Laufzeit selten eine optimale Lösung gegenüber kombinatorischer Optimierungsalgorithmen auf [Ma20]. Dies wird damit begründet, dass es das grundlegende Ziel von RL-Verfahren ist, sowohl schlechte Lösungen zu vermeiden als auch eine durchschnittlich gute Lösung zu erzielen. Folglich wird somit auch das Ziel der Prototypen definiert. Im Gegensatz zu dem Prototyp betrachtet der Google-OR-Solver die Struktur des Problems ganzheitlich und erreicht, mit genug Laufzeit und Rechenleistung, asymptotisch eine optimale Lösung. Neben Google-OR gibt es auch andere Alternativen für das Lösen von kombinatorischen Optimierungsproblemen wie beispielsweise Concorde TSP Solver (https://www.math.uwaterloo.ca/tsp/concorde.html) oder die Dienste von openrouteservice (https://openrouteservice.org/).

4.3 Relevanz für die „Letzte-Meile“-Logistik

Im Bezugspunkt der Relevanz des Prototyps für die „Letzte-Meile“-Logistik und unter der Rücksprache kooperierenden Unternehmen, lassen sich folgende Resultate ableiten. Der Prototyp ist für ein geringes Stoppvolumen von weniger als 1200 Stopps pro Tag nicht effizient genug. Eine der Begründungen ist, dass der manuelle Dispositionsaufwand und das Einsetzen von kombinatorischen Optimierungsalgorithmen für weniger als 400 Stopps in einem kleinem „Letzte-Meile“ Gebiet deutlicher geringer sind als der Aufwand des Trainierens des Prototyps. Erst ab 400 Stopps schwächelt der Google-OR-Solver in der Laufzeit mit 10 s. Unter der Retrospektive und dem Feedback des kooperierenden Unternehmens, wurde die Entscheidung des RL für die „Letzte-Meile“ überdacht und Ideen bezüglich anderer ML-Systeme ausgearbeitet. Ein Favorit ist es Supervised-Learning einzusetzen, um die Dispositionsart der Zulieferer-Disponenten zu lernen, sodass die ML-Komponente unterstützend wirkt und nicht automatisiert VRP-Instanzen löst.

5 Zusammenfassung und Ausblick

Der vorliegende Beitrag bietet ein Konzept zur Basis Implementation eines Reinforcement Learning (RL)-Verfahrens, in Form von „REINFORCE mit Baseline“ kombiniert mit einem Ant Colony Optimization Algorithm (ACO) zum Lösen eines

Vehicle Routing Problems (VRP). Darüber hinaus stellt der entwickelte Prototyp auf Basis der Erweiterungsimplementierung von „REINFORCE mit Baseline“ eine Alternative gegenüber Drittanbietern wie beispielsweise Google-OR da. Des Weiteren werden Impulse zum adaptiven Lösen von VRP mittels verschiedener Optimierungsmechanismen innerhalb des Prototyps geboten. Mit Ausblick auf die weitere Forschungslandschaft von Machine-Learning (ML) und VRP haben die Untersuchungsergebnisse gezeigt, dass eine erweiterte Implementation von RL-Verfahren ein gutes bis sehr gutes Ergebnis im Kontext von einem NP-schweren Problem hinsichtlich der untersuchten Bewertungspunkte und gegenüber den klassischen Optimierungsverfahren erzielen können. Für die weitere Zukunft des Prototyps wäre ein Reduzieren der Komplexität der einstellbaren Parameter denkbar, um mögliches „overfitting“ zu vermeiden. Eine Erweiterbarkeit des Prototyps hinsichtlich anderer RL-Verfahren wäre möglich und bedeutsam, um die Resultate zu validieren. Eine Erkundung weiterer RL-Verfahren ist dahingehend wichtig, da im Gebiet von RL die geringsten Änderungen der Parameter oder des verwendeten Verfahrens starke Abweichungen im resultierenden Ergebnis auslösen. Es wäre in diesem Zusammenhang lohnend zu untersuchen, wie sich das Implementieren von Proximal Policy Optimization (PPO), welches eine neue Art von Policy-Gradienten-Methoden im RL beschreibt, auf die Lösungsqualität auswirkt. Dabei zeichnet sich PPO vor allem durch eine einfachere Implementation und einen einheitlichen Aufbau aus [Sc17].

Literaturverzeichnis

[BRN15] Braekers, K., Ramaekers, K. & Nieuwenhuyse, I.: „The Vehicle Routing Problem: State of the Art Classification and Review“, Computers & Industrial Engineering, Vol. 99, 2015.

[BU21] Bundesverband Paket und Expresslogistik e. V. (BIEK) (Hg.) (2021) KEP-Studie 2021 – Analyse des Marktes in Deutschland: Eine Untersuchung im Auftrag des Bundesverbandes Paket und Expresslogistik e. V. (BIEK), KE-CONSULT Kurte&Esser GbR [Online]. Verfügbar unter https://www.biek.de/download.html?getfile=2897 (Stand: 17.08.2021).

[Ch04] Chang, H. S., W. J. Gutjahr, Yang, J. & Park S.: „An ant system approach to Markov decision processes“, Proceedings of the 2004 American Control Conference, 3820-3825 vol.4, 2004.

[DR56] Dantzig, G. B. & Ramser, J. H.: „The Truck Dispatching Problem“, Management Science, Vol. 6, No. 1, S. 80–91 [Online]. Verfügbar unter http://www.jstor.org/stable/2627477, 1959.

[Go21] Google (2021) Routing Options [Online]. Verfügbar unter https://developers.google.com/optimization/routing/routing_options#local_search_options (Stand: 01.06.2021).

[Lo20] Lorenz, U.: Reinforcement Learning, Springer Berlin Heidelberg, 2020.

[Ma20] Mazyavkina, N., Sviridov, S., Ivanov S. & Burnaev E.: „Reinforcement Learning for Combinatorial Optimization: A Survey“, CoRR, abs/2003.03600, 2020.

[Pu14] Puterman, M. L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming [Online], Wiley. Verfügbar unter https://books.google.de/books?id=VvBjBAAAQBAJ, 2014.

[RA18] Richard, S. S. & Andrew, G. B.: Reinforcement Learning: An Introduction, A Bradford Book, 2018.

[RSC20] Richter, R., Söding, M. & Christmann, G.: Logistik und Mobilität in der Stadt von morgen. Eine Expert*innenstudie über letzte Meile, Sharing-Konzepte und urbane Produktion, 2020.

[SB18] Sutton, R. S. & Barto, A. G.: Reinforcement Learning: An Introduction [Online], The MIT Press. Verfügbar unter http://incompleteideas.net/book/the-book-2nd.html, 2018.

[Sc17] Schulman, J., Wolski, P., Dhariwal, P., Radford, A., & Klimov, O.: „Proximal Policy Optimization Algorithms“, CoRR, abs/1707.06347, 2017.

[Si17] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., van den Driessche, G., Graepel, T. & Hassabis, D.: „Mastering the game of Go without human knowledge“, Nature, Vol. 550, 354, 2017.

[SJM] Stockhausen, P., Johannsen, A., Maurer, R: Konzeption eines Machine-Learning-Verfahrens zum Lösen von Green Vehicle Routing Problemen. Erscheint in Oberweis, A. et al. (Hrsg.): 35. Tagungsband, Arbeitskreis Wirtschaftsinformatik der deutschsprachigen Hochschulen (AKWI), Berlin, 2022

[Su20] Sun, Y., Wang S., Shen, Y., Li, X., Ernst, A. T., & Kirley, M.; Boosting Ant Colony Optimization via Solution Prediction and Machine Learning [Online], 2020.

[VLM19] Vidal, T., Laporte, G. & Matl, P.: „A concise guide to existing and emerging vehicle routing problem variants“, European Journal of Operational Research, Vol. 286, 2019.

Extraktion und Analyse von Schlüsselwörtern in einer Literaturrecherche zu Quantum Computing

Mazlum Copurkuyu [1], Thomas Barton[2]

Abstract: Durch die große Menge an wissenschaftlichen Publikationen, die meist als unstrukturierte Daten vorliegt, nehmen Komplexität und Arbeitsaufwand eines Literature-Review Prozesses stetig zu. Auch im Forschungsgebiet Quantum Computing hat sich die Anzahl wissenschaftlicher Veröffentlichungen in den letzten Jahren stark erhöht. Dieser Beitrag gibt einen Überblick, wie man Text-Mining-Methoden zur Informationsextraktion bei der Literaturrecherche zu Quantum Computing einsetzen kann. Das zentrale Forschungsziel besteht in der Anwendung von Text-Mining zur automatischen Extraktion und Visualisierung von Schlüsselwörtern auf Basis der Abstracts von wissenschaftlichen Publikationen. Dieser Ansatz verwendet zum einen die TF-IDF-Methode und auf der anderen Seite den Word2Vec-Algorithmus, um die automatische Erfassung sowie die Verarbeitung relevanter Literatur zu ermöglichen. Anschließend wird eine visuelle Darstellung der Ergebnisse wie z.B. dynamische Word-Clouds durchgeführt. Aus der Analyse werden Erkenntnisse für den Forschungsbereich Quantum Computing abgeleitet.

Keywords: Literaturrecherche, Text-Mining, Keyword Extraction, TF-IDF, Word2Vec, Quantum Computing

1 Einführung

Eine grundlegende Aufgabe in der Forschung besteht darin, die vorhandene Literatur durch ein Literatur-Review zu identifizieren und zu verstehen, um den Kontext zu ermitteln, weiterführende Forschung zu betreiben sowie die Forschungsgemeinschaft auf dem neuesten Stand zu halten [Ta20]. Zu diesem Zweck ist es zwingend notwendig, die vorhandene Literatur zu einem Forschungsthema zu berücksichtigen. Diese Aufgabe ist jedoch bei der ständig steigenden Anzahl von Veröffentlichungen kaum noch zu bewältigen, und ihre Auswertung gestaltet sich schwierig. Um das Problem der Informationsflut zu lösen, sollte der Prozess der Literaturrecherche strukturiert sein und etablierten Rahmenwerken folgen.

Einer der bekanntesten und weit verbreiteten Vorgehensweise zur Durchführung eines (manuellen) Literatur-Reviews ist das Framework vom Brocke et al. [vo09], das in Abb. 1 dargestellt ist.

[1] Hochschule Worms, Fachbereich Informatik, Erenburgerstr.19, 67549 Worms

[2] Hochschule Worms, Fachbereich Informatik, Erenburgerstr.19, 67549 Worms, barton@hs-worms.de, https://orcid.org/0000-0001-6736-7040

doi:10.18420/inf2022_76

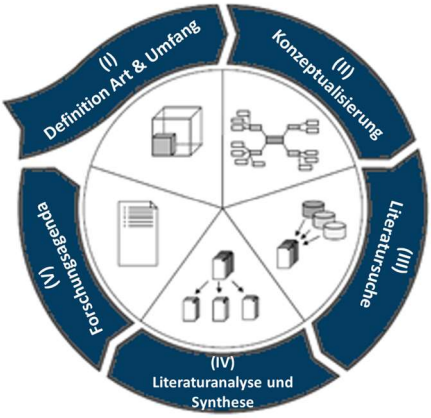


Abb. 1: Prozess zur Durchführung einer Literaturanalyse (vom Brocke et al. [vo09])

Die Durchführung eines Literaturreviews ist jedoch aufgrund der vielen manuellen Tätigkeiten wie Suchen und Herunterladen, Dokumentation des Prozesses, Textscreening usw. sehr zeitaufwändig und mühsam. Das Verfahren zur Durchführung von Literatur-Reviews zu unterstützen und zu automatisieren wäre sehr hilfreich, um mit der immer schnelleren Veröffentlichung von Artikeln Schritt halten zu können [Ta20]. Dabei gilt es zu prüfen, ob die zeitaufwendige Arbeit für die Auswertung sowie die Analyse von wissenschaftlichen Artikeln mit Hilfe von Text-Mining unterstützt werden kann.

In dieser Arbeit werden Methoden des Text-Mining eingesetzt, um Informationen aus wissenschaftlichen Publikationen zu generieren. Hierbei werden Verfahren der Keyword-Extraction und Methoden für Word-Embeddings zur Extraktion von Informationen aus den wissenschaftlichen Abstracts verwendet. Das Ziel ist es, Muster und Verbindungen zu verwandten Themen in einem Forschungsbereich zu ermitteln. Der Beitrag wurde als einer des besten Beiträge der AKWI-Tagung 2022 ausgezeichnet und wird daher im Rahmen dieses Workshops vorgestellt [CB22].

2 Grundlagen

Dieses Kapitel führt in das Quantum Computing und Text-Mining ein. Zuerst wird der Begriff Quantum Computing erläutert. Anschließend wird ein Überblick über Text-Mining gegeben und damit die für diesen Beitrag essenziellen Begriffe sowie Zusammenhänge vermittelt.

2.1 Quantum Computing

Die global vernetzte mobile Welt macht eine nachhaltige und effiziente Verarbeitung und Analyse von Informationen zu einer immer komplexeren Herausforderung. An dem Punkt, an dem klassische Prozessoren heute an ihre physikalischen Grenzen stoßen, könnten Quantencomputer in Zukunft Lösungen liefern. Diese basieren auf der Wechselwirkung quantenmechanischer Zustände [Ba20].

Ein Quantencomputer arbeitet nicht mit klassischen Bits, sondern mit Quantenbits - kurz Qubits. Sie können z.B. durch den Spin eines Elektrons gebildet werden. Die Funktion von Qubits in Quantencomputern basiert auf zwei Schlüsselprinzipien der Quantenphysik: Superposition und Verschränkung, die im Folgenden kurz erläutert werden [Ba20].

Superposition bedeutet, dass ein Qubit nicht nur den Zustand 1 oder 0 einnehmen kann, sondern beide „gleichzeitig“ sowie alle Zustände dazwischen, ähnlich wie eine sich drehende Münze [Ho18]. Solange sie sich bewegt, kann das Ergebnis sowohl Kopf oder Zahl werden. Erst wenn man sie stoppt und den Ausgang misst, wird das Ergebnis bestimmt [Ho18]. Durch die Messung wird dieser Quantenzustand, die Superposition, zerstört. Solange man den Zustand eines Qubits nicht misst, beschreibt man diesen als Superposition, also als Überlagerung aller möglichen Zustände [Ho18]. Der Zustand der Superposition wird durch die Messung und die damit verbundene Interaktion entweder zu 0 oder 1 kollabieren [Ba20].

Die Verschränkung ist der Schlüssel zur Effizienz von Quantencomputern. Verschränkung bedeutet, dass Qubits miteinander unter Wechselwirkungen stehen [Ba20]. Die Verschränkung funktioniert nur, wenn beide Qubits sich in Superposition befinden. Beeinflusst man eines, werden zeitgleich auch alle Partner-Qubits beeinflusst. Es besteht also eine Korrelation zwischen den Zuständen beider Qubits. Sobald der Zustand des einen gemessen wird, ist auch der Zustand des zweiten Qubits bekannt und so kollabiert die Superposition beider Qubits [Ba20]. Dieser Umstand gilt auch bei verschränkten Qubits, die voneinander räumlich getrennt sind. Es sei erwähnt, dass die Verschränkung nach der Messung der Qubits aufgehoben wird [Ho18].

Dadurch ist es möglich, in einem Quantencomputer parallele Berechnungen durchzuführen, aber der Zugriff auf die Ergebnisse der Berechnungen sind eingeschränkt. Der Zugriff auf die Ergebnisse ist gleichbedeutend mit einer Messung, wodurch der Quantenzustand gestört wird. Da die Messung probabilistisch ist, erhält man ein zufälliges Ergebnis [Ba20]. Zudem muss auch die Superposition der Qubits möglichst lange erhalten bleiben, also die Kohärenzzeit möglichst lang sein, denn selbst kleinste Zustandsänderungen verändern den Rechenvorgang. Weiterhin ist die Übertragung der Informationen so empfindlich, dass sie bereits durch das Rauschen der Elektronik gestört werden kann [Ba20]. Solche Fehlerquellen zu minimieren, ist eine spannende Herausforderung.

Die potenziellen Vorteile und Anwendungen des Quantum Computing für die Gesellschaft sind vielfältig. Richtig eingesetzt, sind Quantencomputer unglaublich schnell und effektiv.

Sie können in wenigen Sekunden Berechnungen durchführen, für die die heutigen klassischen Supercomputer viel länger brauchen würden. Diese Tatsache wird von Experten auch als Quantenüberlegenheit bezeichnet [Ba20]. Quantencomputer stellen einen Paradigmenwechsel in der Datenverarbeitung dar. Umso faszinierender ist es, dieses Gebiet zu verfolgen. Daher wird in dieser Arbeit dieses Forschungsgebiet mithilfe von Text-Mining-Verfahren genauer untersucht, um die Themenschwerpunkte, die bisherige Entwicklung sowie die zukünftigen Trends des Forschungsgebietes – sofern möglich – zu erfassen.

2.2 Überblick über Text-Mining

Text-Mining ist ein weit gefasster Oberbegriff, die eine Reihe von unterschiedlichen Verfahren zur Analyse und Verarbeitung von semistrukturierten und unstrukturierten Textdaten beschreiben. Das gemeinsame Ziel hinter jeder dieser Verfahren ist "Texte in Zahlen zu verwandeln", damit leistungsstarke Algorithmen auf große Dokumenten-datenbanken angewendet werden können [Mi12]. Die Konvertierung von Text in ein strukturiertes numerisches Format und die Anwendung analytischer Algorithmen erfordert das Verständnis, wie man diese Verfahren für die Analyse von Texten sowohl nutzt als auch kombiniert [Mi12].

Text-Mining ist ein weitgehend automatisierter Prozess für die Gewinnung von neuen Erkenntnissen aus Textdokumenten [Mi12]. Die Anwendungen des Text-Mining sind so breit gefächert und ihre Ziele so vielfältig, dass es schwierig ist, seine Leistung in allgemeinen Worten zu beschreiben. Im Gegensatz zu anderen etablierten statistischen Methoden ist Text-Mining eine relativ neue und nicht standardisierte analytische Methode zur Wissensentdeckung. Folglich ist es eine Herausforderung, eine allgemeine Vorgehensweise für Text-Mining zu erstellen. Anwendungen des Text-Mining werden in erster Linie durch “trial and error“ auf der Grundlage persönlicher Erfahrungen und Vorlieben bestimmt [Mi12]. Während Data-Mining-Methoden relativ ausgereift sind, gibt es in der Literatur keine allgemein akzeptierten Methoden, die die besten Praktiken im Text-Mining in einem beliebigen Bereich widerspiegeln.

Feldmann [FS07] geht auf funktionaler Ebene davon aus, dass Text-Mining dem allgemeinen Modell einiger klassischer Data-Mining-Anwendungen folgt. Auch Hippner und Rentzmann [HR06] betonen, dass Text-Mining einen ähnlichen Aufbau wie ein klassischer Data-Mining-Prozess aufweist, sich aber in der Datenaufbereitung unterscheidet, weil beim Text Mining eine zusätzliche linguistische Datenaufbereitung erforderlich ist, damit die fehlende Datenstruktur rekonstruiert werden kann.

Die Vorgehensmodelle in der Literatur sind zu allgemein, um sie für einen konkreten Anwendungsfall einzusetzen. Die Prozessmodelle aus der Literatur können aber als Grundgerüst dienen und auf den Anwendungsfall dieser Arbeit übertragen werden (siehe folgenden Abschnitt); wobei die große Bedeutung der Datenvorverarbeitungsphase hervorzuheben ist, da sie eine wichtige Rolle in Text-Mining spielt.

3 Einsatz von Text-Mining

Die Abb. 2 zeigt das Vorgehen für die durchgeführte Studie. Es basiert auf dem Vorgehensmodell von Tauchert et al. [Ta20].

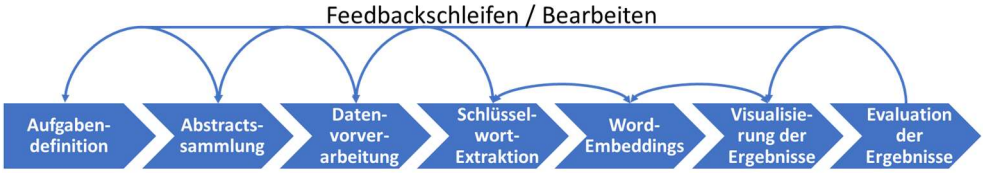


Abb. 2: Grafische Darstellung zur Vorgehensweise bei der Durchführung der Literaturrecherche und -auswertung, die mit Hilfe von Text Mining unterstützt wird [Ta20]

Zunächst werden die Phasen nachstehend kurz erläutert, anschließend werden diese im darauffolgenden Kapitel angewendet. Eine erst kürzlich erschienene Publikation hat die Entwicklung einen digitalen Plattform zum Inhalt, die Forschende dabei unterstützt, eine systematische Literaturanalyse zu unterstützen [Am22].

3.1 Phase 1: Aufgabendefinition

Ähnlich zu Schieber et al. [SH14] und Hippner et al. [HR06] erfolgt eine Festlegung des Zwecks der Recherche und ein Festlegung der Suchbegriffe und der Datenquellen. Um ein gründliches Verständnis zu erlangen und die Ziele genau zu definieren, ist das notwendige Wissen zur Durchführung eines Literatur-Reviews zu ermitteln. Die Definition der Problem- sowie Zielstellung dieser Arbeit ergeben sich aus der Einleitung (siehe Kapitel **Fehler! Verweisquelle konnte nicht gefunden werden.**).

3.2 Phase 2: Abstracts-Sammlung

Die zweite Phase besteht daraus, die wissenschaftlichen Publikationen zu selektieren und die ermittelten Abstracts herunterzuladen, die in den nachfolgenden Prozessschritten analysiert werden sollen. Dazu wird eine Recherche in der englischsprachigen ACM Digital Library [AC22] durchgeführt, und Abstracts zu wissenschaftlichen Publikationen heruntergeladen, die dem vom Benutzer definierten Suchbegriff und dem Datumsbereich entsprechen.

3.3 Phase 3: Aufbereitung der Abstracts mittels NLP-Methoden

In Text-Mining liegen die Textdaten in natürlicher Sprache, d.h. in halbstrukturierter Form, vor. Es ist schwierig, Muster aus halbstrukturierten Daten zu extrahieren, da diese Daten nicht für Text-Mining-Verfahren einsetzbar sind. Um Text-Mining durchzuführen, ist es daher notwendig, die Rohtexte einem Prozess zu unterziehen, bei dem verschiedene Methoden auf sie angewandt werden. Als Ergebnis werden bereinigte Token zurückgegeben. Token sind einzelne Wörter oder Wortgruppen, die als Merkmale für weitere Analysen sowie als Input für Text-Mining-Verfahren dienen [FS07]. Diese verfeinerten Daten sind für den Anwender besser geeignet, um sachkundige Daten zu extrahieren. Dieser Prozess wird als „Vorverarbeitung" von Daten bezeichnet. Die Auswahl der Textvorverarbeitungsverfahren beeinflussen die Darstellung des Textdokuments und damit auch die darauffolgenden Ergebnisse erheblich [FS07]. Viele der Methoden zur Textvorbereitung und -aufbereitung haben ihre Wurzeln in Natural-Language-Processing.

Es wurden drei Vorverarbeitungsmethoden gewählt, wobei die erste die Filterung von Stoppwörtern ist. Dies ist eine einfache und schnelle Methode, um Wörter herauszufiltern, die keine Bedeutung haben und nur in Verbindung mit anderen Wörtern verwendet werden [AHN19]. Das zweite gewählte Verfahren war die Lemmatisierung. Da die eingesetzten Text-Mining-Methoden zur Schlagwortextraktion auf der Häufigkeit des Auftretens eines Wortes in einem Dokument beruhen, ist es notwendig, das Lemma zu identifizieren. Dadurch wird sichergestellt, dass verschiedene morphologische Varianten eines Wortes auf ihr gemeinsames Lemma zurückgeführt und somit als Teil seiner Häufigkeit gezählt werden. In diesem Fall wurde die Lemmatisierung dem Stemming vorgezogen, denn beim Stemming wird einfach ein Wort aufgrund seines Aussehens ausgeschnitten, während die Lemmatisierung ein kalkulierter Prozess ist [AHN19]. Bei der Lemmatisierung ist außerdem zu beachten, dass es oft schwieriger ist, diese für eine neue Sprache durchzuführen als für einen Stemming-Algorithmus. Da eine Lemmatisierung viel mehr Wissen über die Struktur einer Sprache erfordert. Glücklicherweise sind alle Abstracts auf Englisch, so können vorgefertigte Pakete in Python bei der Lemmatisierung verwendet werden. Um gute Ergebnisse zu erzielen, sollten vorher jedoch Tags bezüglich der Wortart mithilfe von POS-Tagging an die Lemmatisierung übergeben werden, da sonst möglicherweise nicht alle Wörter auf die gewünschten Lemmata reduziert werden können [AHN19].

3.4 Phase 4: Keyword-Extraktion

Nachdem alle relevanten Abstracts in einem einheitlichen Format vorliegen, werden im nächsten Schritt die wichtigsten Stichwörter daraus extrahiert. Keywords sind Schlüsselbegriffe, die dabei helfen, die Dokumente zusammenzufassen und einen Überblick über den Inhalt zu vermitteln [FPW99]. Das manuelle Auffinden der relevanten Schlüsselbegriffe, die den Inhalt des Dokuments beschreiben, kann je nach Länge und Anzahl der Dokumente sehr anspruchsvoll sein. Daher wäre die Anwendung eines

automatisierten Prozesses, der Schlüsselwörter aus Dokumenten extrahiert, sinnvoll. Im Bereich des Text-Mining gibt es zahlreiche Methoden der Keyword-Extraction, die z.B. in statistische, überwachte, teilüberwachte und unüberwachte Ansätze kategorisiert werden können [SS15].

Bereits Barton und Kokoev [BK21] nutzten die unüberwachte Lernmethode RAKE, um Schlüsselwörter aus den wissenschaftlichen Arbeiten zu extrahieren. Für diesen Beitrag wird die statistisch-basierte TF-IDF-Methode ausgewählt, da sie Schlüsselwörter für ein bestimmtes Dokument in einer Dokumentensammlung gewichtet. Hierbei werden Wörter als einzigartig (Schlüsselwort) oder unwichtig (Nicht-Schlüsselwort) für das Dokument und somit für den gesamten Korpus eingestuft. Die zu Grunde liegende Idee ist, eine Dokumentensammlung als Beurteilung für die Worthäufigkeit zu verwenden, um die Wichtigkeit eines einzelnen Wortes für ein Dokument festzustellen [MRS08]. Wenn ein Term in vielen Dokumenten vorkommt, geht man davon aus, dass dessen Bedeutung gering ist. Falls ein Begriff in wenigen Artikeln sehr häufig vorkommt, scheint die Relevanz des relevanten Wortes für diese bestimmten Artikel hoch zu sein.

Zunächst berechnet man die Komponente „Termfrequenz“ (TF). Diese bewertet (gewichtet) jedes Wort entsprechend der Vorkommenshäufigkeit. Nun wird ein Blick auf die Formel geworfen. Davor werden zunächst folgende Notationen definiert:

- N ist die Gesamtanzahl der Dokumente
- d ist ein gegebenes Dokument aus dem Datenbestand
- D ist Sammlung aller Dokumente
- w ist ein bestimmtes Word in einem Dokument

TF wird nach der folgenden Formel berechnet

$$\mathrm{tf}(\mathrm{w}, \mathrm{d}) = \log\big(1 + f(w, d)\big),$$

wobei $\mathrm{f}(w, d)$ die reine (Vorkommens) Häufigkeit des Wortes w in Dokument d ist. Die Logarithmierung wurde verwendet, um die Termfrequenz der Dokumente und des Korpus zu normalisieren.

Mit der obigen Formel werden alle Terme als gleich wichtig angesehen. Daher führt man die zweite Komponente „inverse Dokumentenhäufigkeit“ ein, um das Gewicht von Begriffen zu verringern. Die inverse Dokumentenhäufigkeit beschreibt, wie viel Information ein Wort liefert, d. h. ob es in allen Dokumenten häufig oder selten vorkommt. Man definiert die inverse Dokumenthäufigkeit (idf) eines Terms wie folgt [MRS08]:

$$\mathrm{idf}(w, D) = \log\left(\frac{N}{f(w, D)}\right)$$

Man kombiniert nun die Definitionen der Termhäufigkeit und der inversen Dokumenthäufigkeit. Dadurch erhält man das statistische Maß TF-IDF, welches dem Term w eine Gewichtung im Dokument d zuweist, das durch

$$\text{tfidf}(w, d, D) = \text{tf}(w, d) * i\text{df}(w, D)$$

gegeben ist [MRS08].

Diese Arbeit untersucht, wie sich die Themen im Bereich Quantum Computing innerhalb jedes Jahres sowie im Zeitverlauf entwickelt haben. Die Themen werden hierbei als Schlüsselwörter betrachtet und die wichtigsten Begriffe werden auf der Grundlage der Ergebnisse des TF-IDF-Algorithmus pro Jahr identifiziert. Dazu berechnet man den durchschnittlichen TF-IDF-Score aller Wörter über alle Abstracts aus einem bestimmten Jahr.

3.5 Phase 5: Wortrepräsentation mit Word2Vec

Die Methode Worteinbettung (englisch "word embedding") wird in der Computerlinguistik verwendet, um die Bedeutung eines Wortes aus seinem Kontext zu extrahieren bzw. darzustellen. Dieses Verfahren stellt Wörter als numerische Wortvektoren dar, bei denen semantisch ähnliche Wörter auf nahe gelegene Punkte im geometrischen Raum abgebildet werden. Wörter, die in einem ähnlichen Kontext verwendet werden, werden in einem unmittelbaren Vektorraum abgebildet. Das Praktische an der Darstellung von Wörtern als Vektoren ist, dass sie sich dadurch für mathematische Operatoren eignen. Zum Beispiel kann man Vektoren miteinander addieren und subtrahieren.

Es gibt mehrere mögliche Algorithmen, die die Aufgabe der Berechnung von Worteinbettungen lösen können. Angesichts der Beliebtheit und Nützlichkeit wurde für dieses Projekt der Word2Vec-Algorithmus gewählt. Word2Vec ist ein neuronales Netz und kann in Kombination mit der Kosinus-Ähnlichkeitsformel dazu verwendet werden, ähnliche Wörter für ein vorgegebenes Wort zu finden. Das Verfahren wurde im Jahr 2013 von Mikolov et al. [Mi13] bei Google entwickelt. Es sei erwähnt, dass in diesen Beitrag die Funktionsweise sowie die mathematischen Details des word2vec-Modells nicht weiter eingegangen wird. Weiterführende Information sind bei Goldberg et al. [GL14] und Weng [We21] zu finden.

Die Vorgehensweise hier ist es, die Schlüsselwortextraktion zu erweitern. Die Basis-Schlüsselwörter können aus der vorherigen Phase 4 (siehe Abschnitt 3.4) identifiziert werden. Anschließend können weitere ähnliche Schlüsselwörter, die im gleichen Kontext wie die Basis-Schlüsselwörter vorkommen, mithilfe des Word2Vec-Verfahrens sowie dem Kosinus-Ähnlichkeitsmaß ermittelt werden. In diesem Beitrag wird Word2Vec mit demselben Datensatz trainiert, aus dem die vorherigen Schlüsselwörter erstellt wurden. Hierbei dienen die eindeutig identifizierten Schlüsselwörter aus Abschnitt 3.4 3.4als Input, um eine neue Liste von weiteren Schlüsselwörtern zu erhalten.

3.6 Visualisierung der Ergebnisse

Der letzte fehlende Baustein für dieses Projekt ist die Visualisierung der Ergebnisse. Daher werden im Folgenden kurz Informationen über die beiden verwendeten Visualisierungsmethoden bereitgestellt.

Ein Bündel visuell dargestellter Wörter wird als Tag-Cloud, auch Word-Cloud (Wortwolke), bezeichnet [Ku07]. Dieses Bündel von Wörtern wird nach bestimmten Kriterien ausgewählt und hilft bei der explorativen Textanalyse, wichtige Wörter und kontextbezogene Themen aus einem großen Textkorpus hervorzuheben [Ku07]. Im Allgemeinen wird bei der Erstellung von Tag-Clouds die Darstellungsgröße der Wörter herangezogen. Diese ist abhängig von der Häufigkeit eines Wortes. D. h. je häufiger ein bestimmtes Wort in einem Text vorkommt, desto größer erfolgt seine Darstellung in der Tag-Cloud. In dieser Arbeit wird die Häufigkeit durch den berechneten TF-IDF-Wert der Wörter aus der Phase 4 ersetzt und dadurch wird für die Abbildung von Schlüsselwörtern eine relativ aussagekräftige Darstellung ermöglicht.

Für eine Visualisierung der mehrdimensionalen Wort-Vektoren (siehe Abschnitt 3.5) müssen diese in zweidimensionale Wort-Vektoren umgewandelt werden. Dies wird mittels t-Distributed Stochastic Neighbor Embedding (t-SNE) umgesetzt. T-SNE ist ein unüberwachtes und nicht-lineares Verfahren, das von Laurens van der Maatens und Geoffrey Hinton [vH08] im Jahr 2008 entwickelt wurde.

Die Grundidee von t-SNE besteht darin, den mehrdimensionalen Raum unter Beibehaltung des relativen paarweisen Abstands zwischen den Punkten zu reduzieren, wodurch die Abstände zwischen den Punkten identisch bleiben. D. h. der Algorithmus bildet mehrdimensionale Daten auf zwei oder mehr Dimensionen ab, wobei die Punkte, die ursprünglich weit voneinander entfernt waren, auch weit entfernt liegen, und nahe Punkte auch in nahe Punkte umgewandelt werden. Man kann sagen, dass t-SNE nach einer neuen Datendarstellung sucht, bei der die Nachbarschaftsbeziehungen erhalten bleiben.

4 Ergebnisse der Literaturanalyse mittels Text-Mining-Methoden

In diesem Kapitel werden die Ergebnisse der Text-Mining-Anwendung nach der zuvor beschriebenen Methode vorgestellt. Alle Wörter, die mit dem TF-IDF- oder Word2Vec-Verfahren extrahiert wurden, werden bewusst kursiv geschrieben.

4.1 Extraktion von wissenschaftlichen Abstracts

Für diesen Beitrag wird der Suchbegriff „Quantum Computing“ ausgewählt. Um die Suche zu erweitern, wurden die Begriffe „quantum computation“, „quantum computer“ und „quantum computers“ hinzugefügt. Die daraus resultierende Treffermenge wurde durch die Auswahl „Research Article“ in der Kategorie „Content Type“ weiter

eingegrenzt. Für eine bessere Übersicht wurden ausschließlich Artikel von 2008 bis 2021 in Betracht gezogen. Vor dem Jahr 2008 sowie – bedingt durch den Zeitpunkt der Recherche – nach dem Jahr 2021 standen wenige wissenschaftlichen Arbeiten zur Verfügung.

Das folgende Balkendiagramm in Abb. 3 zeigt die Gesamtzahl der Veröffentlichungen für jedes einzelne Jahr. Es ist zu erkennen, dass die Anzahl der Veröffentlichungen von 2008 bis 2021 im Bereich Quantum Computing deutlich gestiegen ist. Im Jahr 2008 wurden in diesem Bereich nur 29 Research-Papers veröffentlicht, während die entsprechende Zahl der Arbeiten im Jahr 2021 196 betrug. Das Wachstum nimmt über die Jahre stetig zu, vor allem ab dem Jahr 2015 ist eine stärkere Zunahme der Veröffentlichungen zu verzeichnen.

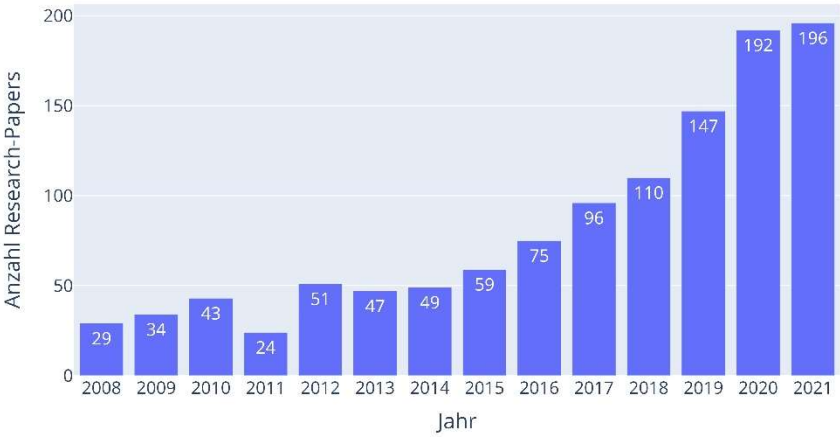


Abb. 3: Entwicklung der Anzahl von wissenschaftlichen Publikationen zu Quantum Computing zwischen 2008 und 2021

4.2 Extraktion der Schlüsselwörter sowie deren Visualisierung mithilfe von Wortwolken

Im Folgenden werden diese Schlüsselwörter ermittelt und visuell analysiert. Hierbei wird die Methode von Weiwei et al. [We10] übernommen. Die vorgeschlagene Methode verwendet die dynamische Tag-Clouds, um die zeitliche Entwicklung von Inhalten zu veranschaulichen. Diese zeitbasierten Word-Clouds stellen verschiedene statische Wortwolken zu konkreten Zeitpunkten dar. Dadurch entsteht ein aussagekräftiges Trenddiagramm, das die Veränderung der Word-Clouds im Laufe der Zeit darstellt. So erhält man einen visuellen Überblick über die unterschiedlichen thematischen Schwerpunkte.

Bei der Erstellung von Word-Clouds werden die durchschnittlichen TF-IDF-Werte herangezogen. Die Größe der Wörter spiegelt den durchschnittlichen TF-IDF-Wert für das

Wort in diesem entsprechenden Jahr wider und gibt dabei die Wichtigkeit für das jeweilige Schlüsselwort – in diesem Fall sind es Bigramme – Abb. 4 zeigt die Wortwolken zu dem beschriebenen Korpus aus dem obigen Abschnitt. Es ist anzumerken, dass Begriffe wie *quantum computing*, *quantum computation*, *quantum computer* und *quantum computers* aus dem Korpus entfernt wurden, da ihre Verwendungshäufigkeit beträchtlich wäre und die anderen Begriffe überdecken würde.

Für die Visualisierung werden zunächst die Abstracts nach Jahr gruppiert und für jedes Jahr wird eine Wortwolke mit jeweils 15 Bigrammen erstellt. Die folgende Abb. 4 zeigt die Ergebnisse für die Jahre von 2019 bis 2021.

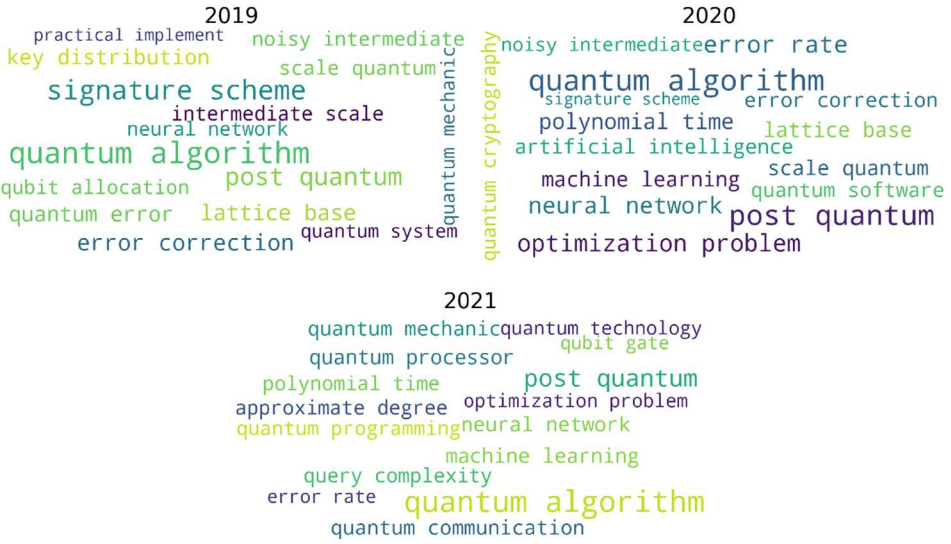


Abb. 4: Visualisierung von extrahierten Schlüsselwörtern mit Hilfe von Word-Clouds für die Jahre 2019 bis 2021

4.3 Zukunft des Quanten Computing

Auf Basis der Wortwolken konnten wichtige Themen für die Zukunft des Quanten Computing identifiziert werden, z. B. *noisy intermediate*, *intermediate scale*, *machine learning*, *artifical intellegence*, *neural network* und *error correction*. Die beiden Bigramme *noisy intermediate* sowie *intermediate scale* deuten auf das offene Kompositum „Noisy Intermediate Scale Quantum Computing“, kurz NISQ.

Im Folgenden wird die zeitliche Relevanz der Wörter untersucht. Für die zeitliche Betrachtung der extrahierten Schlüsselwörter wird eine Kennzahl eingeführt, um eine objektivere Interpretation zu ermöglichen sowie die Bedeutung von Forschungsthemen im

zeitlichen Verlauf abzubilden. Die Kennzahl spiegelt die durchschnittliche Anzahl der Schlüsselwörter pro Artikel für jedes Jahr. Anschließend wird diese Zahl wird für die Erstellung der Liniendiagramme verwendet, wie die folgende Abb. 5 zeigt:

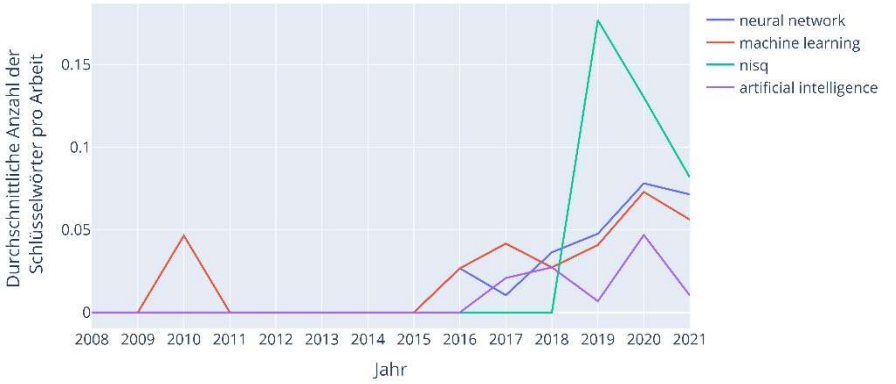


Abb. 5: Zeitliche Entwicklung von Publikationen zu ausgewählten Schlüsselwörtern

Ab dem Jahr 2015 wächst die Bedeutung von Themen wie *nisq*, *machine learning* und a*rtificial intelligence*. Die hohe Anzahl an wissenschaftlichen Publikationen in den letzten drei Jahren (siehe Abb. 3) sowie diese genannten Trends deuten auf eine rasche Ausweitung des Forschungsgebiets Quanten Computing hin und zeigen, dass sich mehr und mehr Wissenschaftler für das Quantum Computing interessierten.

Anhand der identifizierten Stichwörter lässt sich die Zukunft von Quanten Computing in drei Phasen einteilen:

Die erste Phase lautet NISQ. Dies ist die Ära der kleinen, fehleranfälligen oder "noisy intermediate scale quantum" (NISQ) Maschinen, wie Preskill [Pr18] es zum ersten Mal im Jahr 2018 formuliert hat. Verrauscht deshalb, weil man nicht genügend Qubits für die Quantenfehlerkorrektur zur Verfügung hat. Und „*Intermediate Scale*“ wegen ihrer kleinen Qubit-Anzahl.

Nach der NISQ-Ära würde das Rauschen verhindern, dass Quantencomputer mit größerer Anzahl an Qubits eingesetzt werden. Quantencomputer benötigen daher eine Fehlerkorrektur. Ein Ziel wird daher sein, ein verlässliches Verfahren für die Quanten-Fehlerkorrektur in einem Quantencomputer zu implementieren, welches das Rauschen schneller korrigiert als es erzeugt wird. Daher wird solch ein Fortschritt in der Forschung als ein technologischer Meilenstein für die Entwicklung eines fehlertoleranten Quantencomputers gesehen.

Die letzte Phase betrifft vor allem das Wort *machine learning*. Für die nähere Untersuchung der letzten Phase wurde mit dem t-SNE-Algorithmus und dem Word2-Vec-Modell eine zweidimensionale Darstellung der semantisch ähnlichen Wörter im Umfeld des Bigramms *machine learning* durchgeführt. Es werden zunächst die ersten 50

Datenpunkte manuell ausgewertet und anschließend diejenigen mit einer Beschriftung angezeigt, die von Interesse sein könnten. Es sei erwähnt, dass eine Messung der Distanz zwischen zwei Wortvektoren zu keiner gültigen Aussage führt. In der Abb. 6 kann der Abstand zwischen zwei Wortvektoren lediglich als nah oder weit entfernt bezeichnet werden.

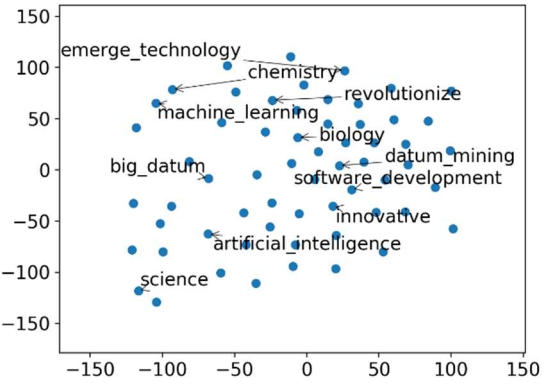


Abb. 6: Zweidimensionale Darstellung semantisch ähnlicher Wörter zu „*machine learning*“

Unter Quanten-Machine-Learning versteht man den Einsatz von Quantencomputern für das maschinelle Lernen. Die in Abb. 6 dargestellten Wörter *innovative*, *emerge technology* sowie *revolutionize* weisen darauf hin, dass Quantum-Machine Learning ein aufstrebendes Gebiet in Wissenschaft und Technologie ist. Es ist der Schnittpunkt von Quantenphysik und maschinellem Lernen. Das maschinelle Lernen könnte mithilfe von Quantencomputern zukünftig in den Bereichen *biology*, *chemistry* und *design automation* eingesetzt werden.

5 Fazit

Im Rahmen einer Literaturrecherche und -analyse zu Quanten Computing lassen sich mit Hilfe von Text-Mining-Methoden Schlüsselwörter identifizieren und auswerten, welche die zeitliche Entwicklung relevanter Themenfelder für das Forschungsgebiet Quanten Computing gut beschreiben. Auch aktuelle Themenfelder, die in einzelnen Zeiträumen vorherrschend waren, lassen sich so ableiten. Die Auswertung einer Literaturanalyse kann mit Hilfe von Text-Mining unterstützt und beschleunigt werden.

Die Anwendung der dynamischen Word-Clouds und der TF-IDF-Methode führten zu dem Ergebnis, dass sich die Forscher/innen in den letzten Jahren verstärkt für die Themen maschinelles Lernen, künstliche Intelligenz, NISQ und Fehlerkorrektur konzentrierten. Es ist zu erwarten, dass diese auch die zukünftigen Trends im Bereich des Quanten Computing bilden werden, da viele Forschungsfragen noch offen sind. Die Einführung

eines zusätzlichen Worteinbettungsverfahren erwies sich als nützlich für das weitere Verständnis der extrahierten Schlüsselwörter und für die Identifikation weiterer Themen und möglicher Trends.

Nach wie vor sind Forschende gefordert, die Literatur zu verstehen, zu verarbeiten und relevantes Wissen zu extrahieren. Insbesondere müssen die Schlüsselwörter, die mithilfe von TF-IDF oder Word2Vec extrahiert werden, genau überprüft werden, da sie nicht immer die eigenen Erwartungen widerspiegeln. So kann beispielsweise eine Word-Cloud ein Schlüsselwort beinhalten, welches zur weiteren Untersuchung weniger geeignet ist. Daher empfiehlt es sich, auf Basis der Word-Clouds oder t-SNE-Visualisierung Stichwörter manuell auszuwählen und diese in einen Wörterkatalog zu übertragen. Darüber hinaus müssen auch Forschungslücken weiterhin eigenständig von Forscher/innen identifizieret werden.

Zusammenfassend lässt sich sagen, dass die Extraktion von Schlüsselwörtern einen unterstützenden Mechanismus zur Beschleunigung des Literature-Review-Prozesses bietet und als ein Hilfsinstrument für die Teilautomatisierung eines manuellen Prozesses betrachtet werden kann.

Literaturverzeichnis

[AC22] ACM Digital Library: ACM Digital Library. https://dl.acm.org/, Abruf am 01.04.2022.

[AHN19] Anandarajan, M.; Hill, C.; Nolan, T.: Practical Text Analytics. Maximizing the Value of Text Data. Springer International Publishing, Cham, 2019.

[Am22] Ammirato, S et al. Digitalising the Systematic Literature Review process: the MySLR platform. Knowledge Management Research & Practice, 2022, S. 1-18.

[Ba20] Bauckhage, C. et al.: Quantum Machine Learning. Eine Analyse zu Kompetenz, Forschung und Anwendung, 2020. https://www.bigdata-ai.fraunhofer.de/de/publikationen/quantum-ml.html. Abruf am 04.04.2022.

[BK21] Barton, T.; Kokoev, A.: Text Mining bei einer wissenschaftlichen Literaturauswertung: Extraktion von Schlüsselwörtern zur Beschreibung von Inhalten. In Barton, T.; Müller, C. (Hrsg.): Data Science anwenden (Angewandte Wirtschaftsinformatik). Springer Vieweg, S. 193–200, 2021.

[CB22] Copurkuyu M, Barton T: Extraktion und Analyse von Schlüsselwörtern in einer Literaturrecherche zu Quantum Computing, AKWI-Tagung 2022, to be published

[FPW99] Frank, E.; Paynter Gordan W.; Witten, I. H.: Domain-Specific Keyphrase Extraction. International Journal of Computer Applications, S. 668–673, 1999.

[FS07] Feldman, R.; Sanger, J.: The text mining handbook. Advanced approaches in analyzing unstructured data. Cambridge University Press, Cambridge, 2007.

[GL14] Goldberg, Y.; Levy, O.: word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method, 2014.

[Ho18] Homeister, M.: Quantum Computing verstehen. Springer Fachmedien Wiesbaden, Wiesbaden, 2018.

[HR06] Hippner, H.; Rentzmann, R.: Text Mining. Informatik-Spektrum 4/29, S. 287–290, 2006.

[Ku07] Kuo, B. Y.-L. et al.: Tag clouds for summarizing web search results. In (Williamson, C. et al. Hrsg.): Proceedings of the 16th international conference on World Wide Web - WWW '07. ACM Press, New York, New York, USA, S. 1203, 2007.

[Mi12] Miner, G. et al.: Practical text mining and statistical analysis for non-structured text data applications. Elsevier Academic Press, Waltham, Amsterdam, Boston, Heidelberg, London, New York, Oxford, Paris, San Diego, San Francisco, Singapore, Sydney, Tokyo, 2012.

[Mi13] Mikolov, T. et al.: Efficient Estimation of Word Representations in Vector Space, 2013.

[MRS08] Manning, C. D.; Raghavan, P.; Schütze, H.: Introduction to information retrieval. Cambridge University Press, New York, 2008.

[Pr18] Preskill, J.: Quantum Computing in the NISQ era and beyond. Quantum 2, S. 79, 2018.

[SH14] Schieber, A.; Hilbert, A.: Entwicklung eines generischen Vorgehensmodells für Text Mining. Technische Universität Dresden, Fakultät Wirtschaftswissenschaften, 2014.

[SS15] Siddiqi, S.; Sharan, A.: Keyword and Keyphrase Extraction Techniques: A Literature Review. International Journal of Computer Applications 2/109, S. 18–23, 2015.

[Ta20] Tauchert, C. et al.: Towards an Integrative Approach for Automated Literature Reviews Using Machine Learning. In (Bui, T. Hrsg.): Proceedings of the 53rd Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences, 2020.

[vH08] van der Maaten, L.; Hinton, G. E.: Visualizing Data using t-SNE. Journal of Machine Learning Research 9, S. 2579–2605, 2008.

[vo09] vom Brocke, J. et al.: Reconstructing the giant: On the importance of rigour in documenting the literature search process: ECIS, 2009.

[We10] Cui, W. et al.: Context-preserving, dynamic word cloud visualization. In: 2010 IEEE Pacific Visualization Symposium (PacificVis). IEEE, 2010. S. 121-128

[We21] Weng, L.: Learning Word Embedding. https://lilianweng.github.io/lil-log/2017/10/15/learning-word-embedding.html, Stand: 19.09.2021.

Concept, Implementation and Evaluation of a Virtual Learning Environment for Acquiring Competences in Android App Development

Frank Neumann [1], Juan Carlos Rodríguez-del-Pino[2], Sebastian Homer[3]

Abstract: The presented project provides a virtual learning environment for university courses teaching app development using Android and Kotlin. It features both self-guided lessons and associated tests as well as practical app development tasks and associated self-assessed programming exercises. The well-known Moodle plug-in *Virtual Programming Lab* (VPL) serves as a starting point for the latter part. VPL gets extended by means for the usage of the Gradle build tool, the Android SDK tools and AndroidX tests for unit tests at the level of Android activities and fragments. In particular, the rather excessive resource requirements of the Gradle tool created numerous problems in the sandboxes used by VPL. These issues were addressed by configuring suitable VPL settings. In addition, problems associated with the proper usage of these settings had to be addressed in VPL. In order to evaluate whether the virtual learning environment suits the student's need for learning app development, a questionnaire was designed that contains questions about the course in general and about the design, handling and the benefits of the exercises The answers of the students to this questionnaire confirm that the proposed virtual learning environment provides a suitable means for learning in the described field.

Keywords: virtual programming lab, VPL, automated grading, automated assessment, automated evaluation, Kotlin, Android, Gradle, Android X tests

1 Introduction

Learning to develop native Android apps is an exciting and challenging endeavor for undergraduate students in computer science. It involves obtaining manifold skills ranging from the used programming language, over various aspects of Android apps and associated frameworks, to high-level architecture and programming concepts needed for such app development.

In the present case, computer science undergraduate students (fifth semester) enroll for the course with a strong background in object-oriented programming, software engineering

1 HTW – Hochschule für Technik und Wirtschaft Berlin, Fachbereich 2: Ingenieurwissenschaften - Technik und Leben, Wilhelminenhofstr. 75a, 12459 Berlin, Germany, frank.neumann@htw-berlin.de, https://orcid.org/0000-0002-8530-3283
2 ULPGC - University of Las Palmas de Gran Canaria, Department of Informatics and Systems, Tafira Campus, 35017, Spain, jc.rodriguezdelpino@ulpgc.es
3 HTW – Hochschule für Technik und Wirtschaft Berlin, Hochschulrechenzentrum, Ostendstraße 25, 12459 Berlin, Germany, sebastian.homer@htw-berlin.de

doi:10.18420/inf2022_77

and some foundations in formal aspects of computer science as well as UML. The learning module starts from this competence level of students and aims at providing them with the above-mentioned theoretical foundations and practical skills for app development up to an intermediate level. The course's goal consists in the development of a medium size Android app by project groups of two students. For mastering the project, the students must take into account modern paradigms of Android development as advocated by the Android team with regards to architecture, design and navigation.

Altogether, the module consists of 12 units that each take three hours (1.5 hours lecture, 1.5 hours of programming exercises and homework) and three additional sessions offering advice and consultation time for the final app project.

The course is split into four learning sections as shown in Table 1.

#	Section name	Number of units	Content
1	Learning Kotlin	5	• Type system • Function • Classes and interfaces • Collections • Lambdas • Coroutines
2	Android basics	2	• Activities and layouts • Views and view groups • Intents • Manifest
3	Android app architecture	4	• Fragments • Navigation • Data binding • MVVM • Connectivity • Asynchronous programming
4	UI aspects	1	• Styles • Themes • Material design • Menus • App bars

Table 1: Content of the Android app development module

The first section focuses on providing insights into the Kotlin programming language and its concepts. In particular, the students learn the type system, functions, classes, interfaces, collection classes, lambdas and coroutines. The subsequent block on Android basics focuses on the various components used in an Android app (activities, intents, manifest, etc.) and layouts. The following section aims at providing architectural guidance for modern Android app development as recommended by Google's architecture guides [1]. The last unit focuses on UI aspects such as themes, styles and menus.

2 Related Work

The following sections analyze what kind of approaches and solutions exist for a virtual learning environment for Android app development using Kotlin. The first section studies the topic of self-guided learning courses and their associated tests for self-assessment. The second section focuses on practical app development tasks and associated self-assessed programming exercises.

The analysis will take into account video guided self-learning courses (e.g. *Udacity*, *Udemy*, *Stepik*), self-learning tutorials (e.g. *Google codelabs*), and some tools aimed at creating self-assessed programming exercises (e.g. *CodeRunner*, *VPL*, *EduTools* plugin). Special attention will be paid to the integration into Learning Management Systems (LMS) used in the university context.

2.1 Self-guided Learning Courses – Conceptual Level

Over the last decade massive open online courses (MOOC) emerged and gained attraction both in academia (e.g. *OpenHPI*[4], *Standford Online*[5]) and on commercial eLearning platforms (e.g. *Coursera*[6], *Linkedin Learning*[7], *Stepik*[8], *Udacity*[9], *Udemy*[10]). MOOCs typically combine video lectures with subsequent assessment of competences. Although many free courses are available, no free MOOCs for the discussed topic of Android app development could be identified. Many of the above-mentioned MOOCs and eLearning platforms provide the necessary tooling to author new learning courses.

Overall, MOOCs are standalone solutions that in most cases are not integrated in the LMS of universities. In the opposite direction, typical LMS (e.g. *Blackboard*, *Canvas*, *ILIAS*, *LAMS*, *Moodle*) used by universities may easily serve as platforms to setup self-guided learning courses combined with automated assessments. Courses in LMS may contain a

4 https://open.hpi.de
5 https://online.stanford.edu/
6 https://www.coursera.org/
7 https://www.linkedin.com/learning/topics/linkedin-learning
8 https://stepik.org
9 https://www.udacity.com/
10 https://www.udemy.com

blend of videos and other learning materials as well as test activities used to assess the learning progress.

Google Developers Codelabs[11] provide self-learning tutorials through a step-by-step experience to setup an app or to use a particular feature in an existing app. They typically give precise coding instructions along with the necessary conceptual knowledge. On the subject of Android app development with Kotlin *Codelabs* offers excellent courses. Unfortunately, *Codelabs* do not provide means for assessing the learning progress, as their main target consists in guiding students through the stepwise process of writing code for the app.

2.2 Self-learning Exercises – Programming Level

A couple of platforms developed over the last decade that support self-learning programming exercises combined with automated assessment. *Codecademy*[12] is a commercial platform offering coding classes in currently 12 programming languages at different levels of proficiency. Users of the platform write their coding solutions, which can then be evaluated. *JetBrains Academy*[13] falls also in this category but offers significant fewer subjects. This platform uses the *EduTools* plugin[14] for Android Studio, IntelliJ and other IDEs. In addition, EduTools may also be connected to other eLearning platforms like Coursera. Unfortunately, no interface connecting *EduTools* to any LMS could be identified. *freeCodeCamp*[15] focuses on providing free self-learning courses initially for web-technologies (HTML, CSS and JavaScript) but extended its scope towards subjects as Python and data science. Unfortunately, these eLearning providers do not integrate in any of the LMS used by universities.

On the university side, various add-ins may be employed to build self-learning programming exercises within LMS. For Moodle two noteworthy plug-ins were identified. The *Virtual Programming Lab* (VPL) [2] [3] [4] [5] [6] [7] [8] is a plug-in to the Moodle LMS, which supports more than 30 programming languages out of the box. It can be easily extended for additional programming languages by supplying adapted bash scripts. It allows teachers to author programming exercises alongside with the expected results as testcases. Students can edit, run and evaluate their solution in the commonly used Moodle environment. If desired, teachers can accept the grades determined by self-evaluation performed by the students. In addition, *VPL* provides teachers with capabilities to detect plagiarism among the handed-in solutions. *CodeRunner* is another plug-in to Moodle, which supports about 10 programming languages but does not offer any check for plagiarism [9].

11 https://codelabs.developers.google.com/
12 https://www.codecademy.com/
13 https://www.jetbrains.com/academy/
14 https://github.com/JetBrains/educational-plugin
15 https://www.freecodecamp.org/

3 Problem Analysis

In the authors perception, the ideal virtual learning environment for university courses teaching app development using Android and Kotlin would allow for the authoring and usage of:

- self-guided lessons including videos and other multimedia learning materials focusing on the teaching of concepts
- associated tests assessing the reached competences of the students
- self-learning programming exercises
- automated assessment of the students' solutions handed-in for the programming-exercises

In addition, the learning environment would allow the teacher to check for plagiarism among the coding solutions. Ideally, the students would work in their usual IDE for Android app development i.e. Android Studio. In order to keep all the data for the course in one place, the results of the tests and programming exercises should be stored within the LMS. Table 2 summarizes the functional and technical requirements identified for the virtual learning environment.

ID	Group	Description
F-1	Authoring content, tests and exercises by teachers	
F-1.1		Authoring of lessons including videos and other multimedia learning materials focusing on the teaching of concepts.
F-1.2		Authoring of tests assessing the competence level of the students in the area of concepts.
F-2	Exploring content by students	
F-2.1		Explore and use course material.
F-2.2		Assessment of acquired competence level based on tests.
F-3	Authoring of programming exercises by teachers	
F-3.1		Authoring of programming exercise description.

F-3.2		Authoring of sample solution for programming exercise.
F-3.3		Coding of test cases used for the assessment of the programming exercise.
F-4	Editing, running and evaluation of coding solutions by students	
F-4.1		Editing of coding solution for programming exercise.
F-4.2		Upload of coding solution for programming exercise.
F-4.3		Running of coding solution for programming exercise.
F-4.4		Evaluation of coding solution for programming exercise.
F-5	Check for plagiarism by the teacher	Identification of similar coding solutions for programming exercises.
T-1	LMS	Moodle should be used as LMS for storing the content and the results of tests and programming exercises.

Table 2: Identified functional and technical requirements for the virtual learning environment

4 Proposed Solution

Starting from the gathered requirements and the technology landscape laid out in the previous sections, suitable technologies will be identified in a first step. Based on the selected technology stack, the desired solution will be further developed and described.

4.1 Selection of the Technology Stack

Chapter 2 described currently available technologies, platforms and tools for the realization of a virtual learning environment geared towards app development for Android and Kotlin. Those technologies will be filtered by the requirements collected in chapter 3 in order to select the best fitting technologies.

Due to the technical requirement *T-1* demanding an LMS integration, the many eLearning platforms unable to integrate with Moodle will be rejected. Moodle itself sufficiently covers the requirements groups *F-1* and *F-2*. Only *CodeRunner* and *VPL* as Moodle plug-ins will be further considered for the coverage of the requirements clusters *F-3*, *F-4* and *F-5*. Neither of both plug-ins supports developing Android apps out of the box. However,

VPL supports the Kotlin programming language and offers plagiarism checks. In contrast, *CodeRunner* does not support Kotlin and does not provide any kind of plagiarism check. Consequently, the combination of Moodle and *VPL* covers more requirements than *CodeRunner* combined with Moodle. Neither of the two solutions supports Android app development out of the box. Finally, Moodle in combination with the *VPL* plug-in is selected as the technology stack to implement the virtual learning environment geared towards app development for Android and Kotlin.

4.2 Configuration of VPL Jail Server

Based on the chosen technology stack of Moodle and *VPL*, the first stage of implementing the desired solution will be detailed. The *VPL* jail server needs to be properly configured for building and running Android apps. In the *VPL* architecture, the jail server acts as an execution sandbox that effectively isolates code run by different users from each other. The development of Android apps requires the following parts of the Android SDK:

- *sdkmanager*: Allows for the installation of the different SDK components.
- *build-tools*: SDK component required for building Android apps.
- *platform-tools*: Contains platform tools like adb.
- *platforms*: The actual SDK packages

On the jail server, the *"/usr/lib/android-sdk"* directory is used for installing the SDK. The SDK manager[16] needs to be downloaded to this directory first. With its help the other three components will be installed. The following code lines depict the installation of the above-mentioned SDK components for the Android SDK version 30.0.3 using the *sdkmanager* tool:

```
sdkmanager "build-tools;30.0.3"
sdkmanager "platform-tools" "platforms;android-30"
```

Subsequently, experiments were conducted on how to use the Gradle build environment within a sandbox on the jail server. The Gradle tool has rather excessive resource requirements that create numerous problems in the sandboxes used by *VPL*. These issues were addressed by identifying and configuring suitable *VPL* settings. It turned out that both within the *VPL* plug-in and on the jail server the maximum used memory and the maximum file size of the execution file had to be increased to 2 GB. Here the required settings for the jail server's configuration file:

```
MAXMEMORY=2Gb
```

16 https://developer.android.com/studio/command-line/sdkmanager

```
MAXFILESIZE=2Gb
```

When using those limits, a jail server of version 2.7.2 or higher is required due to issues in the XMLRPC protocol.

4.3 Configuration of VPL Plug-in

Accordingly, the following limits have to be permitted in the configuration of the VPL plug-in:

```
Maximum memory used: 2 GB

Maximum execution file size: 2 GB

Maximum number of processes: 200
```

4.4 Authoring of VPL Activities

When creating a new VPL activity in Moodle, those settings have to be selected under *Advanced Settings -> Maximum execution resources limits*.

A general issue is the high number of files being used when developing an Android app in Android Studio. The layout of a minimal app consists of various files in a very specific structure depicted in Figure 1.

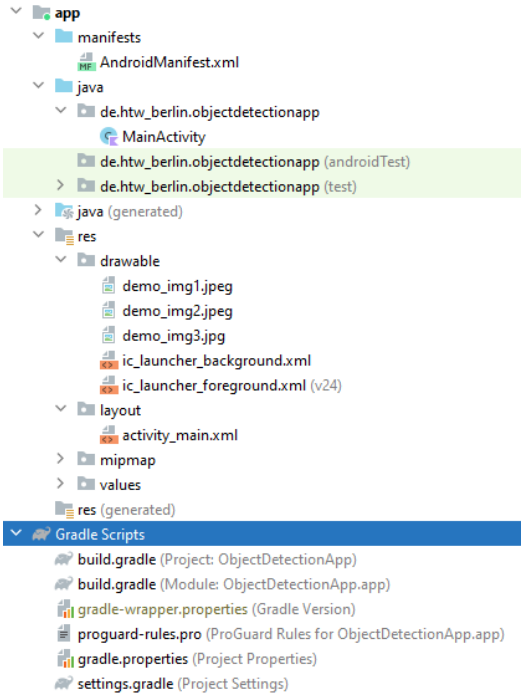


Figure 1: Structure and content of an Android app project

Besides the actual source files, it contains resource files as well as configuration files for Gradle. Drawing from previous experiences, students often fail to hand in a higher number of files. Therefore, the actual submission for the above-mentioned case will consist of only three files:

- MainActivity.kt
- activity_main.xml
- AndroidManifest.xml

Consequently, the other required files have to be provided by the setup of the *VPL activity* as shown in Figure 2. They should be added under *Advanced Settings -> Execution files.*



Figure 2: Android app files contained as VPL execution files

The Gradle version specified in *gradle-wrapper.properties* should be set to 7.3, as previous versions had memory issues in the jail server's sandbox environment.

Finally, the *vpl_run.sh* and the *vpl_evaluate.sh* files need to be adapted for running and evaluating an Android app. Here, the JDK version 15.0.2 is required for running *RoboElectric* tests at the level of activities and fragments. The difference between the run and the evaluate scripts consists only of the different build targets used for Gradle:

- Run: *compileReleaseSources*

- Evaluate: *testDebugUnitTest*

```
. common_script.sh
check_program java
check_program kotlinc
if [ "$1" == "version" ] ; then
       echo "#!/bin/bash" > vpl_execution
       echo "kotlinc -version &> .kotlinc_version" >> vpl_execution
       echo "cat .kotlinc_version | sed 's/.*kotlin/kotlin/'" >> vpl_execution
       chmod +x vpl_execution
       exit
fi

JUNIT4=/usr/share/java/junit4.jar
if [ -f $JUNIT4 ] ; then
       CLASSPATH=$CLASSPATH:$JUNIT4
fi

export CLASSPATH

# search Android SDK
export ANDROID_SDK_ROOT=/usr/lib/android-sdk

# Use JDK 15.0.2
export JAVA_HOME=/usr/lib/jvm/jdk-15.0.2

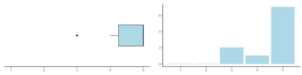
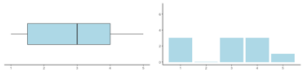
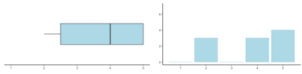
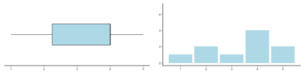
# build and test using Gradle wrapper
chmod +x ./gradlew
cat common_script.sh > vpl_execution
echo "export ANDROID_SDK_ROOT=$ANDROID_SDK_ROOT" >> vpl_execution
```

```
echo "export JAVA_HOME=$JAVA_HOME" >> vpl_execution
echo "./gradlew compileReleaseSources" >> vpl_execution
chmod +x vpl_execution
```

5 Experiences and Discussion

We designed a questionnaire containing questions about the course in general and about the design, handling and the benefits of the exercises. There were 25 students enrolled in the course. 14 of them received grades in the exercises and thus can be regarded as active participants. 10 of the students filled out the questionnaire.

For each question, we used a Likert scale with a range that goes from 1 for the lowest score (strongly disagree) to 5 for the highest (strongly agree). Table 3 contains the questions used, the mean and two distribution plots of the student's responses. The left box plot shows the locality and spread of the answers' numerical values. The median is plotted as a solid vertical line and outliers as dots. The left and right hinges represent the 1. and 3. quartile. The left whisker reaches from the hinge to the lowest value at most 1.5 * inter-quartile range, the right one to the largest value not further than 1.5 * inter-quartile range [10]. The right histogram shows the absolute number of each answer value in the data set.

Question	Mean	Distribution plots
The programming exercises were an important supplement to the lectures.	4.5	
The goal of the tasks provided by the professor was always clear.	3.8	
I had no technical difficulties handing in and evaluating my solutions.	2.9	
The provided exercises made me think about related problems of programming.	3.8	
I was able/asked to create my own solutions for the given problem.	3.4	

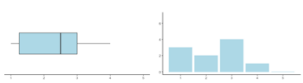
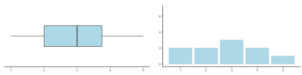
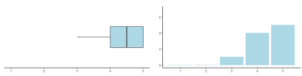
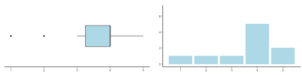
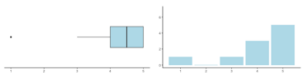
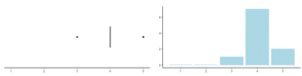
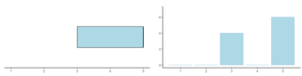
I used other sources to solve the problem.	2.3	
I discussed my approaches with other students in my class.	2.8	
The programming exercises helped me to better understand the concepts of Android apps such as activities, fragments, navigation and data binding.	4.4	
I developed skills and techniques.	3.6	
The lecture and exercises increased my interest in developing Android app.	4.1	
I feel empowered to solve typical problems/issues in developing Android applications.	4.1	
The combination of lectures and practical exercises was a good preparation for the final project	4.2	

Table 3: Evaluation questions and statistics

A pleasant result of the questionnaire – the didactic goal of the course was achieved: Only one student neither agreed nor disagreed with the statement, that the exercises helped to understand the basic concepts of programming Android apps, the others agreed (4) or strongly agreed (5). 7 out of 10 students strongly agreed, that the programming exercises were an important supplement of the lectures. The majority reports an increased interest in developing Android Apps and feel empowered to solve problems or issues in that field.

3 students were faced with technical problems when handing in or evaluating their solutions, only 1 student strongly agreed, that no problems occurred. None of the students asked the technical helpdesk for assistance or failed uploading a solution.

The answers show that students did not use other sources than the provided broadly. We asked them for a list of used sources, they mentioned Google Codelabs, StackOverflow and video tutorials hosted by YouTube.

6 Conclusions

Altogether the proposed virtual learning environment provides a suitable means for learning in the described field. This statement was confirmed by the questionnaires handed in by the students. However, certain aspects of the current setup can still be improved. The answers to the question on technical difficulties when handing in and evaluating solutions suggest that especially the reliability of this aspect can be significantly increased.

Another area of improvement is the long build time required by the Gradle tool. When using Gradle from the command line or from within Android Studio, the long build time is likewise encountered when building a project for the first time. After the first build, Gradle keeps the results of previous builds. On this basis Gradle is able to perform incremental builds that require significantly less time.

However, the sandbox environment of the VPL jail server is not designed to keep results of previous builds. Therefore, Gradle always executes a very time-consuming full build. Improvements addressing this issue have to be implemented at the level of *VPL* and the jail server. Overall, the *VPL* developers are studying how to incorporate the Gradle build tool, the Android SDK tools as well as the AndroidX tests so can be used in *VPL* out of the box.

On the level of the requirements gathered in chapter 3, all of them were completely covered by the selected technology stack and the implemented solution.

Remarks:

The present paper is a slightly extended version of [11].

References

[1] Google, "Android app architecture," 7 10 2021. [Online]. Available: https://developer.android.com/topic/architecture. [Accessed 23 3 2022].

[2] J. C. Rodríguez-del-Pino, E. Rubio Royo and Z. Hernández Figueroa, "A virtual programming lab for Moodle with automatic assessment and anti-plagiarism features," in *Proceedings of The 2012 Internacional Conference on e-Learning, e-Business, Entreprise Information Systems, & e-Government*, 2012.

[3] "VPL, the Virtual Programming lab for Moodle," [Online]. Available: https://vpl.dis.ulpgc.es/. [Accessed 23 3 2022].

[4] A. v. Wangenheim, J. E. Martina, R. Cancian and J. C. Dovicchi, Developing Programming Courses with Moodle and VPL: The Teacher's Guide to the Virtual Programming Lab, Florianopolis: Bookess, 2015.

[5] D. Thiébaut, "Automatic evaluation of computer programs using Moodle's virtual programming lab (VPL) plug-in," *Journal of Computing Sciences in Colleges,* p. 145–151, 01 06 2015.

[6] J. Rodríguez, E. Rubio-Royo and Z. Hernández, "FIGHTING PLAGIARISM: METRICS AND METHODS TO MEASURE AND FIND SIMILARITIES AMONG SOURCE CODE OF COMPUTER PROGRAMS IN VPL," in *3rd International Conference on Education and New Learning Technologies*, Barcelona, Spain, 2011.

[7] J. Rodríguez, E. R. Royo and Z. Hernández, "SCALABLE ARCHITECTURE FOR SECURE EXECUTION AND TEST OF STUDENTS' ASSIGNMENTS IN A VIRTUAL PROGRAMMING LAB," in *3rd International Conference on Education and New Learning Technologies*, Barcelona, Spain, 2011.

[8] J. Rodríguez, E. R. Royo and Z. Hernández, "USES OF VPL," in *5th International Technology, Education and Development Conference*, Valencia, Spain, 2011.

[9] "CodeRunner," [Online]. Available: https://coderunner.org.nz/. [Accessed 3 4 2022].

[10] "A box and whiskers plot (in the style of Tukey)," [Online]. Available: https://github.com/tidyverse/ggplot2/blob/HEAD/R/geom-boxplot.r. [Accessed 3 4 2022].

[11] F. Neumann, J. C. Rodríguez-del-Pino and S. Homer, "Acquiring Android App Development Skills in a Virtual Learning Environment: Extending the Virtual Programming Lab for Moodle towards Self-Assessed App Development Tasks employing Android and Gradle," in *35. Jahrestagung des Arbeitskreises Wirtschaftsinformatik der deutschsprachigen Fachhochschulen (AKWI)*, Berlin/Germany, 2022.

Resilient Smart Farming Lab (RSFLab)

RSF-Lab'22: Resilient Smart Farming Laboratory: Für eine widerstandsfähige und intelligente Landwirtschaft

Christian Reuter[1], Daniel Eberz-Eder[2], Franz Kuntke[2], Matthias Trapp[3]

Abstract: Die zunehmende Vernetzung und Digitalisierung bringen große Veränderungen aber auch Vulnerabilitäten auf allen Ebenen mit sich. Um eine Infrastruktur für ein resilientes Smart Farming (RSF) zu erstellen, welche die Fortschritte der Digitalisierung in der Landwirtschaft nutzt, ohne die Ausfallsicherheit der landwirtschaftlichen Primärproduktion und damit die Lebensmittelversorgung der Verbraucher zu gefährden, bedarf speziell der Sicherheitsaspekt einer kritischen Auseinandersetzung. Der Workshop adressiert diese Forschungsherausforderungen durch Beiträge zu einem umfassenden Monitoring für den Transport von künstlichen Besamungsdosen, zu modularer, sicherer und robuster Steuerungsarchitektur für autonomes Bewirtschaften von Weinbergen, zur Resilienz im ländlichen Raum, zum Aufbau eines informellen Netzwerkes zur Förderung der Digitalisierung in der Landwirtschaft und zu Effizienz und Nachhaltigkeit durch Green-IT.

Keywords: Resilienz, Edge Computing, Landwirtschaft

1 Einleitung

Im Ernährungssektor in Deutschland gilt ein Landwirt als Betreiber kritischer Infrastruktur, sobald ein Schwellenwert der Produktion von 434.500 t Speisen oder 350 Mio. Liter Getränke pro Jahr überschritten wird. Angriffe auf in der Landwirtschaft verbreitete IT-Systeme könnten jedoch viel mehr Betriebe treffen, als in der Verordnung zur Bestimmung kritischer Infrastrukturen durch den Schwellenwert definitorisch erfasst werden und so noch viel massivere Auswirkungen auf die Ernährungssicherheit haben.[4]

Zukünftig soll das *Smart-Farming* durch die Erhebung und Analyse von Prozess- und Sensordaten eine präzise und nachhaltige Bewirtschaftung landwirtschaftlich genutzter Flächen ermöglichen. Aktuell nutzen auf dem Markt verfügbare Dienstleistungen und Produkte vor allem das Cloud-Computing, durch das die betrieblichen Daten – also freiwillig bereitgestellte Betriebsgeheimnisse − extern auf Servern in Rechenzentren statt vor Ort gespeichert werden. Sollte ein Cloud-Anbieter diese Daten für einen illegitimen Zweck verwenden, wäre ein Nachvollzug dessen nicht nur schwer umsetzbar, da es sich um seine eigenen Rechnersysteme handelt, sondern die Nutzung zudem auch schwer sanktionierbar.

[1] Technische Universität Darmstadt (TUDa), Wissenschaft und Technik für Frieden und Sicherheit (PEASEC), Pankratiusstr. 2, 64298 Darmstadt, <nachname>@peasec.tu-darmstadt.de, https://orcid.org/0000-0003-1920-038X, https://orcid.org/0000-0002-7656-5919
[2] Land Rheinland-Pfalz, DLR Rheinhessen-Nahe-Hunsrück, Rüdesheimer Str. 60 - 68, 55545 Bad Kreuznach, daniel.eberz@dlr.rlp.de
[3] RLP AgroScience, Breitenweg 71, 67435 Neustadt an der Weinstraße, matthias.trapp@agroscience.rlp.de
[4] Diese Einleitung basiert auf unserem früheren Beitrag [RSE19]

doi:10.18420/inf2022_78

Ein weiteres Problem in der intelligenten Landwirtschaft stellt die Ausfallsicherheit der Vernetzung dar. Versorgt ein einzelner Anbieter, dessen Service oftmals alle Bereiche eines Betriebes abdeckt mehrere große Betriebe, könnte es im Extremfall zu Produktionsausfällen beziehungsweise Versorgungsengpässen kommen, da bei einem Ausfall des Service im schlimmsten Falle die gesamte Geschäftstätigkeit eines Betriebes stillgelegt werden muss. Dies führt zu einer erhöhten Vulnerabilität des Sektors, auch für absichtlich verursachte Ausfälle durch Cyberangriffe (zum Beispiel Denial-of-Service-Angriffe).

Ein Gegeninstrument stellt die zumindest teilweise Errichtung eigener dezentrale Netzwerke dar (zum Beispiel die „Digitale HofBox“), durch die Programme grundsätzlich auch ohne Internetanbindung nutzbar sind. Solche „Offline-First“-Systeme können zudem die gewohnten Online-Fähigkeiten bieten, um so auch eine optionale Steuerung über das Smartphone zu ermöglichen. Zur Erhöhung der Resilienz ist schließlich ein internes Rechner-zu-Rechner System einer zentralisierten Cloud-Lösung vorzuziehen.

Wie das Eingangsbeispiel zeigt, bringt die exponentiell zunehmende Vernetzung und Digitalisierung große Veränderungen und auch Vulnerabilitäten auf allen Ebenen mit sich, denen speziell in kritischen Infrastrukturen wie der Landwirtschaft zunehmend Aufmerksamkeit gewidmet werden muss. Gerade der Sicherheitsaspekt bedarf einer kritischen Auseinandersetzung, um die Infrastruktur für ein resilientes Smart Farming (RSF) zu erstellen, welche die Fortschritte der Digitalisierung in der Landwirtschaft nutzbar macht, ohne die Ausfallsicherheit der landwirtschaftlichen Primärproduktion und damit die Lebensmittelversorgung der Verbraucher zu gefährden.

Aber auch positive Aspekte von digitalisierten Prozessen müssen Beachtung finden und weiterhin untersucht werden. Beispielsweise können breitflächig installierte Sensoren wichtige Daten für zukünftige Krisenprävention liefern, beispielsweise in Form präziserer Frühwarnung bei Extremwetterereignissen. Auch der Aufbau eines eigenständigen Kommunikationskanals für Krisensituationen durch Adaption von Sensornetzwerktechnik ist denkbar und könnte im Ernstfall eine wichtige Ergänzung zur Kommunikation in Krisensituationen sein, wenn andere Kommunikationskanäle nicht mehr zur Verfügung stehen.

2 Aktuelle Beiträge zum Thema

Um Forschung zu Resilient Smart Farming zu fördern und sichtbar zu machen wurde ein Aufruf gestartet, wissenschaftliche Beiträge zu diesem Thema einzureichen:

- Umweltsensoren und Umweltinformatik
- Umweltmodellierung und -simulation
- Erdbeobachtung und Überwachung
- Anwendungen von geographischen Informationssystemen
- Umweltinformationssysteme
- Technologien für eine widerstandsfähige digitale Infrastruktur

- Digitale Technologien und Robotik in der Landwirtschaft
- Rationales Naturmanagement und Ökologie in der landwirtschaftlichen Produktion
- Smart Farming und Precision Farming
- Semantic Web und Ontologien im Zusammenhang mit der Landwirtschaft
- Cybersicherheit in kritischen Infrastrukturen

Die auf Basis eines Peer-Reviews selektierten Beiträge adressieren dieses Thema in vielfältiger Weise.

Die Autoren Paul Schulze, Frank Fuchs-Kittowski (Hochschule für Technik und Wirtschaft Berlin), Tim Hafemeister (Institute for Reproduction of Farm Animals Schönow), Alexander Urban, Mario Berndl, Christian Simmet (Minitüb GmbH) und Martin Schulze (Institute for Reproduction of Farm Animals Schönow) untersuchen in ihrem Beitrag „*Development of a measuring system for monitoring transport of boar semen from artificial insemination centers to sow farms*" die Anforderungen an ein systematisches Vorgehen zur Erfassung aller relevanten Umweltbelastungsfaktoren beim Transport von Besamungsdosen basierend auf einer Befragung von Besamungszentren. Mit dem vorgestellten mobilen Messsystem wird ein umfassendes Monitoring für den Transport von künstlichen Besamungsdosen möglich.

In ihrem Beitrag „*A modular control architecture for safe and robust robot operation and inspection in steep slope vineyards*" stellen Eike Gassen, Patrick Wolf und Karsten Berns (Robotics Research Lab, TU Kaiserslautern) eine modulare Steuerungsarchitektur für sicheres und robustes autonomes Bewirtschaften von Weinbergen in Steilhangumgebungen vor. Tests in einem authentischen Weingebiet in der Nähe der Mosel in Deutschland beweisen die Machbarkeit und Robustheit des Ansatzes.

Der Beitrag „*AgriRegio: Infrastruktur zur Förderung von digitaler Resilienz und Klimaresilienz im ländlichen Raum am Beispiel der Pilotregion Nahe-Donnersberg*" des Forschungsprojektes AgriRegio von Christian Reuter, Franz Kuntke (TU Darmstadt), Matthias Trapp (RLP AgroScience), Christian Wied (IBM Deutschland GmbH), Gerwin Brill (expeer GmbH), Georg Müller (Maschinen- und Betriebshilfsring Rheinhessen-Nahe-Donnersberg e.V.), Enno Steinbrink, Jonas Franken (TU Darmstadt), Daniel Eberz-Eder (DLR Rheinhessen-Nahe-Hunsrück) und Wolfgang Schneider (Maschinen- und Betriebshilfsring Rheinhessen-Nahe-Donnersberg e.V.) stellt vor, wie die digitalisierte Datenerfassung und -nutzung in landwirtschaftlichen Betrieben widerstandsfähiger gemacht und die sicherheitskritische Infrastruktur geschützt werden können. Sieben Projektpartner erproben dazu smarte Sensoren auf Basis standardisierter Open-Source-Technologien in der Landwirtschaft, bei denen die Betriebsdaten dezentral auf lokalen Servern gespeichert werden.

Der Beitrag „*Aufbau eines informellen Netzwerkes zur Förderung der Digitalisierung in der Landwirtschaft in Rheinland-Pfalz*" von Lea Wintz und Paul Strerath (Technische Hochschule Bingen) untersucht im Rahmen des Verbundprojektes Experimentierfeld Südwest des Bundeslandwirtschaftsministeriums in Rheinland-Pfalz die Bedarfe der landwirt-

schaftlichen Betriebe in der Digitalisierung, um mittels Coaching und Schulung die Digitalisierung in landwirtschaftlichen Betrieben weiterzuentwickeln. Über ein herstellerunabhängiges, zentral gesteuertes informelles Netzwerk der Hochschule für angewandte Wissenschaft, wird der kritische Dialog zwischen Anwendern gefördert.

Im Beitrag „*Effizienz und Nachhaltigkeit durch Green-IT: ein systematischer Literaturüberblick im Kontext der Klimakrise"* stellen Laura Buhlmeier, Patrick Gantner, Tobias Frey, Michael Bohrs, Mar-André Kaufhold und Christian Reuter (TU Darmstadt, Wissenschaft und Technik für Frieden und Sicherheit (PEASEC)) die Ergebnisse einer systematischen Literaturrecherche vor, in der Herausforderungen wie der hohe Energie- und Materialverbrauch der IKT-Geräte und Datenzentren sowie Potenziale, beispielsweise durch Effizienzsteigerungen, in der Bewältigung der Klimakrise durch eine effiziente und nachhaltige Entwicklung des IKT-Sektors analysiert werden.

Literaturverzeichnis

[RSE19] Reuter, C., Schneider, W., & Eberz, D. (2019). Resilient Smart Farming (RSF) – Nutzung digitaler Technologien in krisensicherer Infrastruktur, *39. GIL-Jahrestagung: Informatik in der Land-, Forst- und Ernährungswirtschaft Fokus; Digitalisierung für landwirtschaftliche Betriebe in kleinstrukturierten Regionen – ein Widerspruch in sich?*, Lecture Notes in Informatics (LNI), A. Meyer-Aurich (Hrsg.), S. 177-182, Vienna, Austria: Gesellschaft für Informatik. https://gil-net.de/Publikationen/139_177.pdf.

Development of a measuring system for monitoring transport of boar semen from artificial insemination centers to sow farms

Paul Schulze [1], Frank Fuchs-Kittowski [2], Tim Hafemeister[3], Alexander Urban[4], Mario Berndl[4], Christian Simmet[4] and Martin Schulze[5]

Abstract: During transport of boar semen for artificial insemination of pigs, there is currently no mobile measuring system available for recording all relevant impact factors systematically. Based on a survey of artificial insemination centers, this paper describes the requirements for a systematic approach to record all relevant environmental impact factors during the transport of artificial insemination doses. The developed system architecture as well as the implementation to fulfil the derived requirements is presented. An Arduino Nano 33 BLE Sense with an external temperature sensor and an external luminance sensor is used as the measuring device. This device sends measurement data to a smartphone app using Bluetooth Low Energy (BLE). The app persists the measurement data on a local storage on the smartphone. With the presented mobile measuring system, a comprehensive monitoring for transport of artificial insemination doses becomes available.

Keywords: Data acquisition, data logger, system architecture, measurement system, sperm transport, boar

1 Introduction

Boar semen is a highly sensitive product, which requires special demands to maintain quality [Ms17]. Previous studies shows an influence of transport on sperm quality [Sc18]. In today's pig reproduction liquid preserved boar semen is delivered from artificial insemination centers to downstream sow farms using standard passenger cars or small trucks. Relevant vibrations occur during transport of artificial insemination doses, which might have a negative impact on semen quality [Ht22]. There are other impact factors which occur during transport as well e.g., temperature fluctuations. With existing systems (data loggers, etc.) both factors can only be recorded insufficient. The literature shows contributions where mobile measuring systems are used to record vibrations [FSA17, SDW18,

[1] Hochschule für Technik und Wirtschaft, Fachbereich 2, Umweltinformatik, Wilhelminenhofstraße 75A, 12459 Berlin, Germany, Paul.Schulze@htw-berlin.de, https://orcid.org/0000-0003-4997-7931
[2] Hochschule für Technik und Wirtschaft, Fachbereich 2, Umweltinformatik, Wilhelminenhofstraße 75A, 12459 Berlin, Germany, Frank.Fuchs-Kittowski@htw-berlin.de, https://orcid.org/0000-0002-5445-3764
[3] Institute for Reproduction of Farm Animals Schönow, 16321 Bernau, Germany, t.hafemeister@ifn-schoenow.de
[4] Minitüb GmbH, Hauptstraße 41, 84184 Tiefenbach, Germany, minitube@minitube.de
[5] Institute for Reproduction of Farm Animals Schönow, 16321 Bernau, Germany, m.schulze@ifn-schoenow.de

doi:10.18420/inf2022_79

SLS19]. However, the solutions presented here have been developed only for specific applications. There is no system for a mobile recording of the impact factors during the transport of boar semen available. Therefore, the development of an adapted measuring system is necessary.

The aim of this paper is to show development, implementation and evaluation of a mobile measuring system for the systematically recording of impact factors during the transport of artificial insemination doses. The mobile measuring system will be integrated as a component in a real-time monitoring system for the transport of boar semen.

This paper is structured as followed: Systematic approach and methods used are presented in section 2. Relevant factors which have an impact on transport stress of boar semen and which should be recorded by our proposed system are shown in section 3. This is followed by a requirements analysis for the planned measurement system. Based on these results a general architecture and then a technical concept for the measurement system are presented. Section 4 describes implementation of the measurement system. In section 0 results are discussed. The paper closes with a summary and an outlook in section 6.

2 Methodical approach

A literature review was done to determine all relevant impact factors on transport stress of boar semen, which affect quality during transport. Furthermore, a qualitative survey of six selected breeding companies (production companies for insemination doses of boar semen) with a total of about 10,000 boars in Brazil, Germany and the USA was performed as part of the research project "IQ-TranS"[6]. The aim of the survey was to analyse the current state of transport process as well as supply chains. For this purpose, an interview guide was developed that takes into account the current situation in delivery process of artificial insemination doses. Then a stakeholder analysis was done to identify relevant stakeholder for the mobile measurement system. As a result of the survey requirements for a measurement system were determined and a system architecture was designed that fulfil these requirements. Based on this concept, the mobile measuring system was implemented. Finally, the mobile measuring system was validated in a field test.

3 Concept

In the following, the concept of the measurement system is presented. For this purpose, the results of the literature review and the expert interviews are summarized. Furthermore, requirements for the measurement system are determined. The general architecture and the technical concept for the implementation of these requirements are presented afterwards.

[6] Website: www.iqtrans-projekt.de

3.1 Relevant impact factors on transport stress of boar semen

The nature of boar semen demands high standards for transport of this perishable product. Numerous influencing factors are known, such as the dilution technique, storage, **tem perature**, **light**, etc., which affect the quality (negatively) [HGK16, Ms17]. Besides these factors, rotation of the semen doses as well as the transport itself can have a negative influence on quality of boar semen as well. Recent studies shows that **vibrations** during transport can have a negative influence on the quality of boar semen too [Ht22, Sc18]. Reduction in quality of semen doses leads to a reduction in economic value. [FBS18, SRW15]. The Literature review shows that boar sperm is a highly sensitive product that requires very specific demands during transport (constant cold chain at 16 °C and low-emission transport) in order to maintain its condition and quality.

3.2 Requirements analysis

The evaluation of the expert survey with focus on the transport process showed, that there is a need for documentation and tracking of supply chains. Semen doses are transported from artificial insemination center to the customer by land over distances of up to 500 km and for up to 12 h. The transport process is very heterogeneous among the companies surveyed and is summarized below. Tubes or blisters are usually used as transport packaging for boar semen doses. Packaged semen doses are transported in actively temperature-controlled boxes, insulated polystyrene boxes with warm or cold packs, or without special insulating packaging within a temperature-controlled cargo hold. Logging of the temperature of the transported goods takes place irregularly. Standard passenger cars or small trucks are used as transport vehicles. Shipment is usually carried out by instructed employees or commissioned subcontractors. In-house employees were trained in handling of vulnerable semen doses. Interviewed breeding companies have a high interest that all relevant impact factors on transport stress of boar semen are recorded with an automatic system and stored for later analysis.

For the planned mobile measurement system, the following stakeholders were identified: the artificial insemination center and the delivery persons. The artificial insemination center needs measurement data, which are collected during transports. These transports are carried out by the delivering persons. The aim of the planned mobile measuring system is to automatically record all relevant impact factors which have an influence on transport stress of boar semen. The mobile measuring system is part of a real-time monitoring system in prospect for the digitalization of boar semen transport (see: [Sp22]).

In order to fulfil the requirements of the artificial insemination centers surveyed, the planned measurement system must be embedded into the existing delivery process. The requirements and use cases for the measuring system are described below:

- All relevant impact factors on transport stress of boar semen during the entire transport process have to be recorded. These are temperature inside the transport box used or temperature of the cargo hold of the vehicle, light intensity to which the sperm doses are being exposed and vibrations applied to semen doses during the transport process.
- For documentation and tracking of the transport chain, the location should be continuously recorded during the entire transport process.
- The measuring system is planned to be used in delivery vehicles. It should be possible to install it into or upgrade existing transport boxes. For this purpose, sufficient miniaturization, battery operation and wireless data transmission have to be taken into account.
- All measurement data have to be stored for analysis and evaluation after measurement. For this purpose, a standardized, memory-saving and human-readable data format has to be used.
- Quality requirements have to be fulfilled, too. This includes fault tolerance of the measuring system: No failure of the measuring system in case of a failure of individual components as well as the interchangeability of components used.

3.3 General architecture

Due to the required hardware, the general architecture of the planned measurement system cannot be considered separately from the hardware. Therefore, the planned hardware is described below in abstract form as far as possible and specified in the following section 3.4. below.

Sensors are required to measure environmental parameters as well as fulfil performance requirements. The following sensors types are required to measure the environmental parameters: accelerometer, thermometer, luxmeter, GNSS receiver. The measurement of acceleration should take place with at least 50 Hz (see: [FBS18]). All other environmental parameters must be measured at least every 5 s. The runtime of a single recording is at least 12 h.

To fulfil the identified requirements a general architecture was developed. The measuring system is divided into several parts. A **programmable datalogger** is used to query and control the **sensors** used as well as to transmit the measured environment parameters. An **app** is used to control the data logger. The measured values are received from the data logger by the app and persisted in the **data storage** of the smartphone used. Additionally, the app is used to determinate the current **location** using a GNSS receiver in the smartphone. Figure 1 shows the general architecture of the measurement system.

Commands and measured values are as transmitted via a text-based (non-binary) key-value data model. The advantages of a key-value data model are very good scalability, high performance and easy modification [SK20]. This leads to easier data transfer between the subsystems and a faster implementation. [BA20].

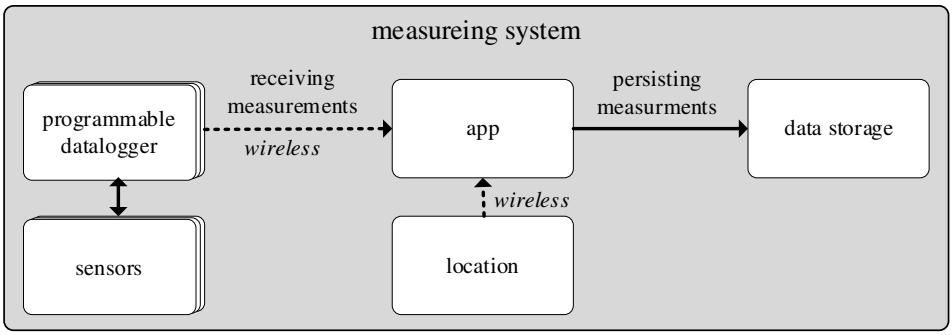


Figure 1: General architecture of the measurement system

To meet the quality requirements of fault tolerance and interchangeability, only hardware components with standardized interfaces should be used. Furthermore, the data transfer between the subsystems should be realized with existing and stable interfaces.

3.4 Technical concept

Based on the general architecture and the requirements analysis, the required hardware components for the measurement system are determined and described below. The following standards can be used for wireless data transmission: Bluetooth, Bluetooth Low Energy (BLE), WiFi and WiFi Direct [CGS13, GOP12]. Radio technologies such as LoraWAN and ZigBee are not suitable due to their limited data rate [JSM22].

Numerous programmable microcontrollers are available that can be used as data loggers [JSM22]. For intended use the "Arduino Nano 33 BLE Sense" is suitable. This microcontroller has built-in sensors such as: accelerometer, thermometer, barometer, luxmeter and more. The board has a size of approx. 45 mm x 18 mm. With the decision to use this microcontroller, the radio interface was set to BLE or WiFi by hardware [Ar22]. Compared to WiFi BLE has the advantage that no access point is required inside the vehicle and the overhead for data transmission is lower, which enables a faster implementation [JSM22]. For this reason, BLE is used to transmit the measured values from the microcontroller to the app. By using existing interfaces (like I²C, SPI, ...) additional sensors and peripheral devices can be connected to and operated by the microcontroller as well [Ar22].

Own examinations showed that the integrated temperature sensor of the Arduino Nano 33 board is influenced by the other components on this board and has a temperature offset of at least 5 °C. Therefore, an external 10 kΩ thermistors (temperature dependent resistor) is used to measure the temperature of insemination doses.

The external luminance sensor "Seeed Studio Grove APDS9002" is used to measure light intensity. All sensors read out by the microcontroller are directly connected to it or already integrated on the board. To determine the location by GNSS, a smartphone running the

data logger app is used. All Measured data is persisted on the file storage of the smartphone used. The microcontroller is powered by an integrated battery circuit with a Li-Ion battery and can also be charged via this circuit.

The selected components are common and well available standard components with a sufficient degree of miniaturization. All components of the measuring system fitted together fit well in one "hand". Figure 2 shows the component diagram of the measuring system with the sensors and modules used.

4 Implementation

The implementation of the measurement system is described below. Firmware was written in C for the microcontroller (data logger), which queries all sensors with adjustable sampling rates and then transmits the measured values via BLE. The measuring rate for the acceleration sensor was set to 50 Hz and for the other sensors to 0.2 Hz. The microcontroller publishes a GATT service as a server for data transmission and registers a “characteristic” for each environmental parameter. These characteristics are used to published the measured environmental parameters. Each microcontroller can be uniquely identified via its MAC address. The GATT characteristics are uniquely assigned using previously defined UUIDs. Figure 3 shows the definition of the GATT service for data transmission via BLE. Due to the GATT protocol, each measured parameter is sent in a separate message. Data transmission is based on the push principle i.e., the receiver (client, the app) registers for the characteristics to be received. Transmission of the measured values takes place in "notify" mode and is therefore independent of the receiver i.e., there is no delay in receiving an acknowledgement of receipt [HP20]. The data logger stores the measured values temporarily until transmission. Persistence is not implemented on the data logger.

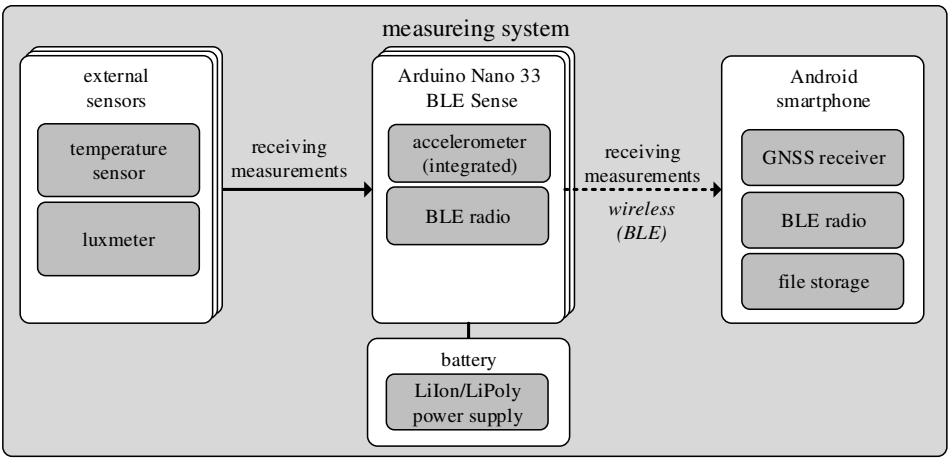


Figure 2: Component diagram of the measuring system with the sensors and modules used

An Android app was written in Java as receiver. The app searches for BLE devices in the surrounding area and can connect to several data loggers simultaneously to receive measurement data from these devices at the same time. After initiating the connection, the latest measurement data is transmitted directly to the app. Due to the use of Bluetooth Low-Energy, it is possible to establish several connections between measuring devices and the smartphone at the same time. A simultaneous transmission of measurement data from several measuring devices was successfully accomplished. However, it should be noted that wireless radio is always a shared medium which leads to a limitation of data transmission. Especially at high measurement rates, as our system has, upscaling is limited. To eliminate this bottleneck, in the future measurement values for acceleration will be calculated on the microcontroller (Arduino Nano) and summarized to a displacement index. The displacement represents the average vector length between two sampling points per second (see equation 1 in [Ht22]). This reduces our transmitted data rate for acceleration to 1 Hz. Therefore, a sufficient scaling is possible to connect all transport boxes of a vehicle to the smartphone.

The app receives the measured values and persist them as a text-based (non-binary) data format on the smartphones integrated file store system. Thereby the measured values are stored in a key-value based data format. For each key-value pair the measurement time (Unixtime in millisecond) and an ID of the measuring device are stored as well. This allows each measurement to be uniquely assigned.

A suitable case was produced for the data logger prototype using a 3D printing process. This case protects the hardware components from moisture and dust and allows an easy handling. To increase the durability of the measuring device further, it is planned to protect the electronic components with epoxy resin. We expect a sufficient protection for the rough environment in the field. In Figure 4 the prototype is shown.

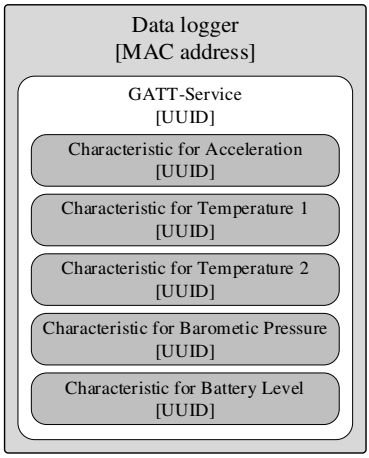


Figure 3: Definition of the GATT service and the included GATT characteristics for publication of the measured values from the data logger via Bluetooth Low Energy (BLE).

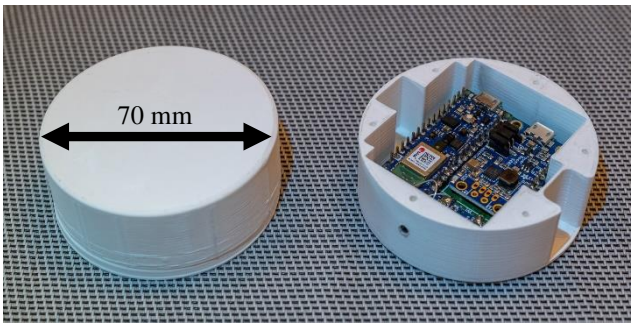


Figure 4: Prototype of a datalogger from the measurement system.

5 Evaluation

For the evaluation of the measuring system, measurement trips from producers (artificial insemination centers) to customers (sow farms) were carried out during a reference delivery with length of about 800 km. The roads covered had different surfaces: smooth asphalt, interrupted asphalt, cobblestone and unpaved dirt road. The measuring system could be successfully tested and validated to determinate vibration immissions during the transport of boar semen (see: [Ht22]). In addition to the vibration immissions, the temperature and the location were successfully recorded during the test trips as well. Our Experience with the system shows that a transport of up to 12 h can be recorded entirely and reliable. For further evaluation of our measurement system, additional test drives are planned with potential users. Handling and reliability will be tested particularly in this field test.

The key-value data format used to store the measurement data, ensured the measurement data can to be easily imported and extended by any other sensors and metadata if required. In the current development state of the mobile measuring system, no efforts have been made to protect the data from manipulation. For a successful deployment under real production conditions, further adjustments are necessary by the development team.

Due to the size of the system components and the integrated power supply our measuring device can either be installed in the transport boxes or it can be placed in a transport box as an external logger (cf. Figure 4) [Ar22]. The use of low-cost common components enables production even in large quantities.

During the measurement trips done, it could be confirmed, that the developed measurement system meets the system quality requirements. Thus, a failure of a sensor does not lead to a complete failure of the measuring system. Even interference in the Bluetooth connection leads only to a temporarily interrupted data transmission. By using the "nofity" mode for data transmission, an interrupted connection is restored as quickly as possible without user interaction. However, data packets that are not received are irreversibly lost and cannot be recovered. Our own tests using "indicate" mode in Bluetooth Low Energy

for data transfer have shown, that data transfer rate drops significantly compared to "notifiy" mode.

With the developed and presented mobile measuring system all the requirements from the interviewed artificial insemination centers are fulfilled (cf. section 3.2 Requirements analysis).

6 Summary and Outlook

In this paper a mobile measuring system was designed, which allows a complete recording and storage of all relevant impact factors (see: [HGK16, Ms17].) during transport of boar insemination doses from artificial insemination centers to downstream sow farms. The measurement system consists of an “Arduino Nano 33 BLE Sense” as a microcontroller with integrated sensors such as an accelerometer as well as other integrated sensors. An external 10 kΩ thermistor was connected to the microcontroller board to determine the temperature of the insemination doses and a “Seeed Studio Grove APDS9002 luminance sensor” was connected to determine the luminance in the transport box. The measurement data is transmitted via Bluetooth Low Energy to a smartphone and our developed app persists the data on the smartphone's integrated memory. The location is determined via the GNSS receiver integrated in the smartphone. It is possible for multiple loggers to be connected to the app at the same time and take measurements simultaneously.

With the presented measuring system, a part of the real-time monitoring system for the transport of boar semen could be implemented (see: [Sp22]). Following research will be focused on the determination of tolerable vibrations (threshold values) for the transport for boar semen. The collected measurement data from the transport trips will be analysed later in the app and suitably presented in a user-friendly front end. For this purpose, a driving assistance app is planned. The app carries out an evaluation of influence on transported boar sperm to determined its current condition (e.g.: green, yellow and red transport conditions). As soon as one of the monitored values (temperature and vibrations) exceeds the threshold value, the user will be warned.

Acknowledgements

This study was based on a project funded with a by the Federal Ministry for Economic Affairs and Climate Action (BMWK) on the basis of a decision by the German Bundestag. Funding reference: 16KN077342 (IQ-TranS)

Bibliography

[Ar22] Arduino: Arduino® Nano 33 BLE Sense - Product Reference Manual.

[BA20] Breedlove, T.W.; Albert, R.L.: C++: An Active Learning Approach: An Active Learning Approach: Jones & Bartlett Publishers, 2010 — ISBN 978-1-4496-6216-5.

[CGS13] Camps-Mur, D.; Garcia-Saavedra, A.; Serrano, P.: Device-to-device communications with Wi-Fi Direct: overview and experimentation. In: IEEE Wireless Communications Magazine Bd. 20, 2013.

[FSA17] Feldbusch, A.; Sadegh-Azar, H.; Agne, P.: Vibration analysis using mobile devices (smartphones or tablets). In: Procedia Engineering, *X International Conference on Structural Dynamics, EURODYN 2017*. Bd. 199, S. 2790–2795, 2017.

[FBS18] Fuchs-Kittowski, F.; Bortfeldt, R.; Schulze, M.: TransportLog 1.0 - Mobile Sensing-App zur Analyse des Einflusses von Transportstress auf die Eberspermaqualität: Gesellschaft für Informatik e.V., 2018. — Accepted: 2019-05-28T11:24:40ZISSN: 1617-5468https://gil-net.de/Publikationen/30_83.pdf — ISBN 978-3-88579-672-5.

[GOP12] Gomez, C.; Oller, J.; Paradells, J.: Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology. In: Sensors Bd. 12, Nr. 9, S. 11734–11753, 2012.

[HGK16] Hoy, S.; Gauly, M.; Krieter, J.: Nutztierhaltung und -hygiene. Stuttgart: Eugen Ulmer KG, 2016 — ISBN 978-3-8252-4369-2.

[HP20] Hoddie, P.; Prader, L.: IoT Development for ESP32 and ESP8266 with JavaScript: A Practical Guide to XS and the Moddable SDK: Apress, 2020 — ISBN 978-1-4842-5070-9.

[Ht22] Hafemeister, T.; Schulze, P.; Bortfeldt, R.; Simmet, C.; Jung, M.; Fuchs-Kittowski, F.; Schulze, M.: Boar Semen Shipping for Artificial Insemination: Current Status and Analysis of Transport Conditions with a Major Focus on Vibration Emissions. In: Animals Bd. 12, Nr. 10, S. 1331, 2022.

[JSM22] James, A.; Seth, A.; Mukhopadhyay, S.C.: IoT System Design: Project Based Approach, Smart Sensors, Measurement and Instrumentation. Bd. 41. Cham: Springer International Publishing, 2022 — ISBN 978-3-030-85862-9.

[Sc18] Schulze, M.; Bortfeldt, R.; Schäfer, J.; Jung, M.; Fuchs-Kittowski, F.: Effect of vibration emissions during shipping of artificial insemination doses on boar semen quality. In: Animal Reproduction Science Bd. 192, S. 328–334, 2018.

[SRW15] Schulze, M.; Rüdiger, K.; Waberski, D.: Rotation of Boar Semen Doses During Storage Affects Sperm Quality. In: Reproduction in Domestic Animals Bd. 50, 2015.

[Sp22] Schulze, P.; Fuchs-Kittowski, F.; Hafemeister, T.; Schulze, M.: Digitalisierung des Spermatransports – Anforderungen und Softwarearchitektur: Gesellschaft für Informatik e.V., 2022. — Accepted: 2022-02-24T13:34:51Z — ISBN 978-3-88579-711-1.

[SDW18] Shrestha, A.; Dang, J.; Wang, X.: Development of a smart-device-based vibration-measurement system: Effectiveness examination and application cases to existing structure. In: Structural Control and Health Monitoring Bd. 25, Nr. 3, S. e2120, 2018.

[SLS19] Sony, S.; Laventure, S.; Sadhu, A.: A literature review of next-generation smart sensing technology in structural health monitoring. In: Structural Control and Health Monitoring Bd. 26, Nr. 3, S. e2321, 2019.

[Ms17] Schulze, M.; Schäfer, J.; Jung, M.; Waberski, D.: Reproduktionsforschung – Neue Wege zur Sicherung der Spermaqualität in der Schweinebesamung. In: Züchtungskunde Bd. 89, S. 22–28, 2017.

[SK20] Steven, M.; Klünder, T.: Big Data: Anwendung und Nutzungspotenziale in der Produktion: Kohlhammer Verlag, 2020 — ISBN 978-3-17-036477-6.

A modular control architecture for safe and robust robot operation and inspection in steep slope vineyards[1]

Eike Gassen, Patrick Wolf, Karsten Berns[2]

Abstract: In fall line cultivation of vineyards, the vines are planted vertically to the hill's slope. Therefore, steep slope vineyards require a high amount of manual labor, doubling production costs. Working in such environments is exhausting and laborious. Therefore, autonomous robots should assist humans in reducing costs and increasing safety. However, current state-of-the-art robotic systems and control architectures are not designed to work in such harsh environments with extreme terrains. Therefore, this work proposes and modular control architecture for safe and robust autonomous working in steep slope environments. Tests in an authentic vineyard near the Moselle river in Germany prove the approach's feasibility and robustness.

Keywords: Off-Road Robotics, Steep Slope, Agricultural Robots, Behavior-Based Control, Control Architectures

1 Introduction

A vineyard's layout is typically in fall line cultivation. In this kind of cultivation, the vines are planted vertically to the hill's slope. The authors of [SML21] highlight that extremely steep slope vineyards suffer from the disadvantage of higher manual labor, resulting in more than twice the labor costs per hectare considered over one working year. However, there is a cultural desire to preserve the cultivation of these steep slopes, which have a decisive influence on the appearance of wine-growing regions such as the Moselle or Ahr valleys in Germany. From an ergonomic point of view, working an entire day in such an environment is exhausting and laborious for a person. Therefore, there is a particular need to automate this steep slope work. According to a Geisenheim University survey of 500 self-marketing wineries with steep slopes [SML21], the median winery has an area of 2.5 ha of steep slope vineyards. This small amount often does not justify purchasing a crewed steep slope vehicle, whose acquisition costs with attachments and associated trailer can reach a lower six-figure amount, depending on the manufacturer. Therefore, an objective is offering a crewless vehicle that is affordable in its acquisition costs even for smaller wine estates and saves additional labor costs through autonomous work.

[1] This work was founded by the Federal Ministry of Food and Agriculture, Germany (BMEL) and is part of the project *experimental field southwest* (`https://ef-sw.de`).

[2] Robotics Research Lab, Department of Computer Science, TU Kaiserslautern, {gassen,patrick.wolf,berns}@cs.uni-kl.de

doi:10.18420/inf2022_80

A particular challenge for autonomous working in steep slopes is the extreme environmental conditions. Typical state-of-the-art algorithms regarding off-road navigation and working consider moderate terrains, e.g., [Br13], [Ga19]. Accordingly, navigation planning and perception need to consider the terrain-based challenges in addition. However, a vineyard is typically clearly structured. Therefore, some assumptions regarding the surroundings can apply, which reduces complexity. Therefore, this work presents a modularized control framework that considers steep slope vineyards' specialties for autonomous traveling and inspection.

This paper introduces configurable, modular control framework for safe and robust inspection of steep slope vineyards. Therefore, Sect. 2 presents related work to autonomous working in vineyard applications and related control architectures. Details on the demonstrator robot are available in Sect. 3. Sect. 4 provides an overview to the integrated behavior-based control architecture iB2C, which is fundamental for designing a robust, fault-tolerant off-road control concept. Sect. 5 presents details of control architecture. There was a special focus on properties as extensibility, robustness, maintainability, and a lightweight implementation. Finally, Sect. 6 provides experiments and Sect. 7 summarizes and concludes.

2 Related Work

There is an enormous variety of agricultural vehicles. The authors of [KFP17] illustrate that machines that work on an agricultural terrain with a slope of more than 60% require a winch to prevent accidents. Therefore, the relevant category for steep slope viticulture vehicles have such a device, and the following focuses on such systems in more detail.

Established viticulture vehicle systems usually consist of two subsystems: A transport platform that moves on public roads, typically a trailer or sometimes a self-propelled tool. The second component is the working vehicle itself driving between the trellis. A person sitting on top controls the vehicle and manually works and switches on or off the corresponding device. Most of these vehicles are crawlers because of their excellent ground contact and the low center of gravity[3]. The drawback of tracked vehicles is the height, weight, and ground erosion caused by the weight.

Existing control technology in agriculture which could be beneficial for the use cases in viticulture, are primarily developed for flat terrain. The authors of [Ad17] show an approach of a teleoperated agricultural robot that applies a spraying agent to the plants. [Sa18] shows the path planning for a robot used in vineyards for crop monitoring. The automated task of dispersion of pheromone dispenser is explained in [Ro18]. There are few crewless, autonomously operating vehicles for steep slope viticulture as Geisi [SBK12], [Br17], or Slopehelper [Ba21].

[3] Exemplary systems exist here https://www.geier.it/de/modelle/ueberblick/

There are two main drawbacks regarding the existing robots: massive construction and limited autonomy skills. In particular, SLOPEHELPER[4] is a large robot with a weight of nearly two tons and a width of 160 cm. Accordingly, the vehicle is too large to safely operate within the narrow vineyard conditions addressed by this contribution. Also, the soil corrosion and ground compaction are too extensive. In general, the autonomy skills of these robots are minimal (e.g., tactile driving to keep the lane). However, there is a demand for automated mapping and documentation through robots to support winemakers and allow intelligent farm management. Accordingly, the robot requires mapping plants, detecting diseases, and estimating harvest failures. Therefore, this work proposes a modular architecture beyond low-level vineyard navigation.

3 Steep Slope Vineyard Inspection Robot

Fundamental design considerations of the applied steep slope inspection robot provides [KGB21]. The robot is a lightweight carrier system with a low center of gravity and can equip various attachments as sprayers (see also Fig. 1). Accordingly, there is a separation of concerns between driving and working subsystems. The robot secures through a steel cable, which controls the longitudinal motion through a winch and prevents falling. Tab. 1 shows the most important driving relevant properties. They define the configuration of the low-level safety systems as stop distance, rollover prediction, or trajectory correction.

Tab. 1: Hardware and kinematic specification of the demonstration robot.

Property	Value	Property	Value
Wheel Base	1.18 m	Max. Velocity	$1.50\,\mathrm{m\,s^{-1}}$
Track Base	0.69 m	Max. Curvature	$0.84\,\mathrm{m^{-1}}$
Vehicle Length	1.71 m	Min. Turning Radius	1.18 m
Vehicle Width	0.76 m	Min. Turning Cycle	2.36 m
Vehicle Height	0.75 m	Max. Ground Clearance	0.26 m
Tire Radius	0.28 m	Tilt Angle*	32.80°
Length KC To Front	0.31 m	Max. Fwd Incl. Angle*	00.00°
Length KC To Back	1.40 m	Max. Bwd Incl. Angle*	46.00°

* under the assumption of a mounted sprayer attachement and higher center of gravity

The driving kinematics of the robot is a non-standard differential drive. Three primary criteria affect navigation planning:

controllable winch the winch secures the robot in steep slopes through a steel cable. Additionally, the cable supports the robot's longitudinal motion and aids the localization systems. The winch is located in front of the robot. Therefore, the robot reverses the inclination and drives forward during upwards climbing.

[4] https://slopehelper.com/parameters

kinematic center The robot has four wheels, but only the (larger) front-facing (to the winch) tires are actively controllable. The rear-facing wheels are passive. Therefore, the kinematic center locates at the front axle on the ground.

active front tires The robot's front tires actively control and correct the motion. During descending the slope, the primary motion vector defines through the surface. Therefore, the motion adjustment occurs only through the front tires.

The inspection robot has a symmetrical sensor layout with four stereo cameras covering each side of the vehicle. Additionally, localization sensors exist. Fig. 1 and Tab. 2 illustrate the sensor setup. For cost reduction, the setup avoids deploying laser sensors. However, with an expected price reduction in the future, this constraint may change.

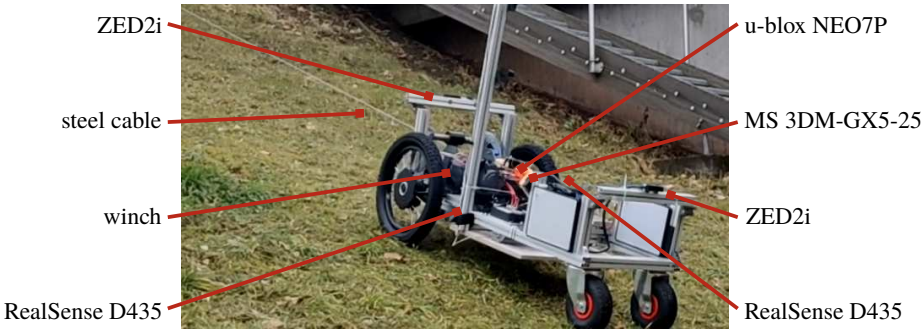


Fig. 1: Early prototype of vineyard inspection robot with the proposed sensor layout during steep slope trials. The robot is secured by a steel cable and controls the inclination climb (here approximately 35°) through a winch. The tires actively correct the robot's motion.

Tab. 2: Corresponding sensor setup. Sensor poses are relative to the kinematic center of the vehicle. (f/ l/ r/ b) indicates a front, left, right, or back mounting. The winch directs to the front.

Sensor	x [m]	y [m]	z [m]	Φ [°]	ψ [°]	φ [°]
ZED2i (f)	1.26	0.00	0.60	0.00°	0.00°	0.00°
ZED2i (b)	-0.32	0.00	0.60	0.00°	0.00°	180.00°
Realsense D435 (l)	0.50	0.34	0.60	0.00°	0.00°	90.00°
Realsense D435 (r)	0.50	-0.34	0.60	0.00°	0.00°	-90.00°
MS 3DM-GX5-25	0.00	0.00	0.60	0.00°	0.00°	0.00°
u-blox NEO7P	0.00	0.00	0.80	0.00°	0.00°	0.00°

4 Integrated Behavior-Based Control

Behavior-based systems have been well known since the 1980s for robots' reactive and robust control. Brook's subsumption architecture changed design of control systems fundamentally [Br86] and many behavior-based robots followed over time, e.g., [Ar98], [Ma90], [Jo04].

A more recent behavior-motivated approach, inspired from the gaming industry and in particular the popular game Halo 2 [Is05], are behavior trees [Io20]. They allow for dynamically reacting to unstructured environments due to the high modularity and reactivity.

The integrated behavior-based control (iB2C) architecture [RWB17], developed at the Robotics Research Lab of the TU Kaiserslautern, provides a behavior-based framework that decouples control and data flows. It includes a partial activity of behavior nodes. Accordingly, it can represent non-discrete states. Therefore, a BBS has a unique way of system arbitration which creates robustness and flexibility. It consists of highly distributed and parallel units with overlapping functionality. The resulting behavior emerges from the direct interaction of those components. Also, iB2C explicitly focuses on modeling behavior-based perception systems.

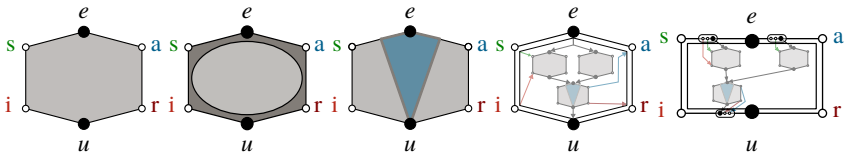


Fig. 2: The different components of the iB2C architecture [Wo21]: Module, Percept, Fusion, Group, Layer.

An important concept of iB2C is modularization, abstraction, and interaction of different components (Fig. 2). Therefore, an iB2C behavior network has multiple (simple) behaviors that interact. Accordingly, a behavior network shows the control and data flows in a graph-like appearance similar to the execution order. A module is the standard unit for control context and percepts tailored towards perception systems. Fusion behaviors resolve conflicts, which arise from parallelism and overlapping functionality. Behavior systems tend to grow large through modularity. Therefore, groups and layers provide encapsulation to avoid arbitrary connections between subnetworks. Therefore, a group acts as a single behavior within the network and denotes a strict encapsulation. Layers encapsulate softly concerning the semantical meaning. Therefore, they provide additional layer interfaces which provide access to selected behaviors encapsulated by the layer.

Each behavior offers a standardized interface for system arbitration using an *Activity Function* $f_a e \in 0,1$. Additionally, data processing uses a specialized interface with the *Transfer Function* Fe that computes output data u based on inputs e. It can be *stimulated* by another behavior's activity via the stimulation input s and *inhibited* via the inhibition input i. The behavior's internal *potential* $\phi = \min\left(s, 1-i\right) \in 0,1$ describes the effective relevance of a behavior in the network. It limits the behavior's *activity* $a = \min\left(\iota, f_a e\right) \in 0,1$ which represents the amount of influence in the current system state. The *target rating* $r = f_a e \in 0,1$ resembles the behavior's contentment with the current system state. While the so called *meta-signals* s, i, a, r are strictly defined in- and outputs, the vectors e and u can carry arbitrary sensor and control data.

An example for a rollover prevention safety system shows Fig. 3. Data flow connections

show black, stimulation links green, and inhibitions red. The network monitors the roll and pitch angles of the robot stops if a critical value is exceeded.

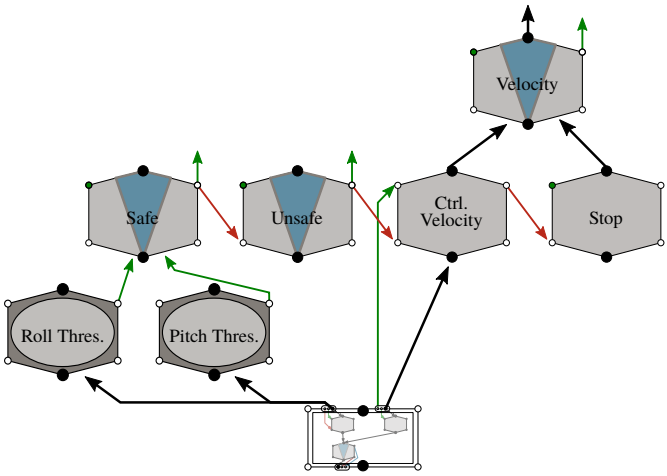


Fig. 3: Example behavior network for rollover prevention in steep slope environments. The robot is default unsafe and inhibits the control velocity. Percept behaviors monitor the safety-relevant inclination thresholds of the robot. If the inclination is within the safety margins, the network actively suppresses the unsafe state. Accordingly, the robot is allowed to set the control velocity. The conflict between stopping and driving resolves the velocity fusion.

5 Modular Robot Control Architecture for Steep Slope Driving

The architecture follows an action-/perception-oriented design [SPB08] and realizes through the behavior-based architecture iB2C. Important properties are

Modularity, extensibility, and maintainability The framework consists of modular control and perception nodes that structure in semantically corresponding layers. The communication realizes by clearly defined interfaces between the layers. This modularity allows for the easy exchange of methods and software components. Accordingly, the system is easily maintainable since the separation of concerns and encapsulation prevent side effects on the remaining software concept. The action-/perception-oriented design further decouples perception and control by exchanging so-called virtual sensors. Therefore, the architecture is also easily extensible since the application requires adding novel hardware for specific tasks to the robot, such as spraying or inspection devices. The corresponding software units must also be easy to integrate, based on the plug-and-play principle.

Robustness The robot operates in a semi-structured environment, but the environmental conditions are still demanding. As the name already indicates, a steep slope vineyard

has extreme inclinations, which the robot must consider during driving to avoid damages or accidents. Therefore, the system must behave robustly to environmental disturbances as the weather, illumination changes, and season-related appearance changes. Further, smaller vegetation such as grass, which is no actual obstacle, is not allowed to affect the robot's control and navigation abilities negatively.

Real-time performance The robot must safely navigate the steep slope environment. Therefore, the control systems require to react in real-time (e.g., 20 ms) to avoid collisions or rollover based on the environment. A particular challenge is the detection of negative obstacles. It is only possible to close the gap due to obstructions, and therefore short-term safety must secure the robot.

Light weight implementation A particular aim is cost reduction for the base robot, primarily a carrier system for attachments or personnel. Therefore, fewer computational resources are available, demanding careful management of resources and, thus, algorithms.

An architecture that features many of these properties is REACTiON [Wo18], tailored towards robust off-road navigation in cluttered environments. However, the default architecture requires high computational resources since semantic scene interpretation, and terrain-based navigation planning is done [Wo20].

In contrast, the terrain where the robot operates is semi-structured, and therefore simplifications regarding scene interpretation are possible. A vineyard contains multiple, clearly structured rows. The general structure of a vineyard is somewhat similar compared to other locations, and the vine lines define the track for navigation, and the robots navigate between them. The steep slope vineyards have their trellis scaffolding and vines aligned with the slope. Therefore, the inclination and the plants define the main driving direction. The robot must follow the primary structures and avoid colliding with them or object within the vineyard.

The proposed control architecture (Fig. 4) focuses on the described properties. It consists of an hardware abstraction layer (blue), perception (yellow), control (red) and user interface (gray). An important property of the layered design are the interface definitions. Accordingly, each layer has a clear role within the framework and implementation are exchangeable.

Hardware Interface (blue) The hardware interface encapsulates the physical hardware, either the robot, a simulation, or data playback of recorded data. Accordingly, the interface provides similar data with similar quality. Therefore, the behavior framework, which accesses the hardware interface, cannot distinguish between real, recorded, and simulation data. This modeling allows for testing the control framework under various conditions. E.g., the simulation allows for testing safety-critical situations, which otherwise would destroy the robot or its surroundings. Therefore, the overall probability of the system causing fatal

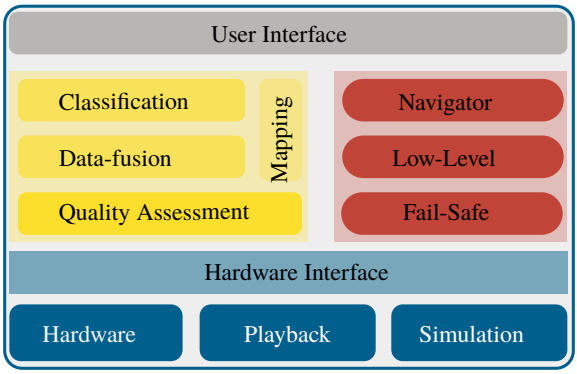


Fig. 4: Overview on the modularized architecture. Each component refers to an iB2C layer and allows a separation of concerns for a high extensibility and maintainability.

damage to the environment or robot reduces. The interface contains the robot's motion data: velocity, curvature, and the steel cable's length. The sensor data streams (camera image, point cloud, IMU, GNSS) are also available.

Perception (yellow) The perception system has different central components: quality assessment, data fusion, classification, and mapping. Initially, the data quality of each sensor system evaluates for disturbances and filters low-quality information [RWB17], [WB21]. Then, data fusion combines different data streams to higher-level information. Examples are point cloud merging and localization filters. Subsequently, an evaluation of data follows. The surrounding categorizes into drivable and non-drivable regions. These segmentation data maps into local grid representations. With this, maps act as short-term memories for sensor data and guide navigation planning.

Control (red) The primary control units are the navigator, low-level, and fail-safe. The navigation planning system is relatively simple due to the structuredness of the vineyard. Nonetheless, a user can advise the robot to navigate to a specific plant or a desired coordinate within the row. A low-level controller monitors the robot's safety by considering obstacles and adjusting the steering to drive collision-free within a row. A fail-safe system avoids falling over (see also Sect. 4).

6 Experiments

Field tests were made in Zell (Moselle), Germany, one of the biggest wine-growing centers in the Moselle wine region. A flat terrain was chosen for initial tests because of the facilitated working conditions. The distance between the vine rows, which is the tramline, is



Fig. 5: Data recording for validation of the control and perception concepts. The sensor setup used on the robot is mounted to an experimental design to check the sensing quality. The Velodyne HDL32E laser serves as the ground truth for the stereo cameras.



Fig. 6: Stereo camera views (front, left, right). The front camera records the steel cable. The sideward facing cameras scan the vine.

equally 2.20 m. In the trellis itself, the vines are planted with a distance of 1.10 m. The first experiments for the sensor setup are done with a frame pulled through a vineyard on a skid (see Fig. 5). Between the first and the second row, the skid was placed and pulled along the row by a winch. On October 22nd 2021, one day before the grape harvest, records were made to receive images of the foliage wall under vintage conditions.

The setup used an additional 3D lidar for reference and ground truth purposes. Otherwise, the sensors matched Sect. 3: The front-facing ZED2i stereo camera served for path planning and the recognition of the vines. The sensor skid contained a RealSense D435 camera on each flank, directed at the foliage wall or the vine. A GPS and IMU sensor was mounted at

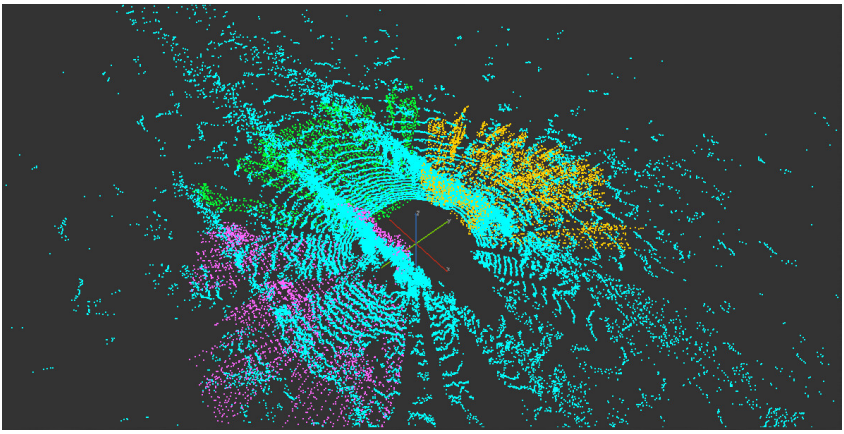


Fig. 7: The combined stereo point cloud. The laser-based reference point cloud is depicted in turquoise. The stereo cameras visualize in green, pink, and yellow.

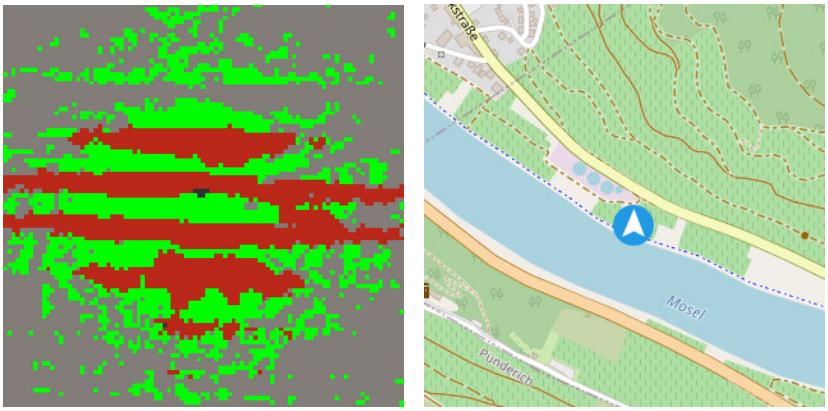


Fig. 8: Occupancy map and localization results. Traversable regions between the vine are labeled green. Plants and objects show in red.

the kinematic center of the vehicle. The ZED camera's roll angle was titled horizontally to investigate the influence of path planning and to create a map.

The skid started at the end of the row. Then, it was pulled with continuous speed along the trellis by a winch. Records of the driven path and the foilage of both sides- and the vine itself were done simultaneously. The orientation and height of the different cameras have been changed during other runs.

Figs. 6–8 visualizes the results of the vineyard tests. An initial observation is that the cameras capture all relevant parts of the vine and the track.

The stereo point clouds match accordingly and build a complete surround view, which is

required to segment the vine from the ground. Also, the main structures (poles, stems) are well separated in the point data. Quality assessment filters distorted points, which are typically farther away. The stereo point clouds match accordingly and build a complete surround view, which is required to segment the vine from the ground. The lookahead distance is between 3–5 m (with a deviation of max 0.1 m), which is sufficient for path planning and feature extraction. The segmented information is stored using occupancy grids, representing free passages between the vine. The navigation systems adjust to the surrounding obstacles. Also, the localization showed a good performance since there are few obstructions in a vineyard, and the steel cable supports localization.

One additional finding was that it is sufficient to use one stereo camera in the vehicle's driving direction to generate path planning and mapping images. Thus, no laser point cloud is required from the quality aspect.

7 Conclusion

The contribution introduced a modular control architecture for autonomous inspection and working in steep slope vineyards. Maintaining the work in steep slope vineyards is a cultural desire but comes at a high cost and is labor intensive. State-of-the-art robot systems (Sect. 2) operate either in flat terrain, are heavy, or do not have a substantial amount of autonomy besides simple (sometimes tactile) lane following. Accordingly, Sect. 3 presents a concept for a steep slope robot, which is lightweight, secured by a winch, and can mount attachments. The vehicle has a sensor setup containing localization sensors and stereo cameras to perceive the conditions of the vines and autonomously operate in the vineyard. Sects. 4–5 provided an overview to behavior-based robotics, in particular iB2C, which implements the modularized robot control architecture. Essential properties are robustness, modularity, extensibility, and efficiency. The framework provides quality assessment, data fusion, classification, and mapping systems. Further, there exist navigation planning and low-level/ fail-safe safety systems. Experimentations occurred in a vineyard near the Moselle river using the described sensor setup slide (Sect. 6). There was a high performance, and the sensors proved to perceive and map the surroundings robustly. The robot successfully localized itself within the vineyard and created maps of the trellis lines.

Future work addresses additional tests with the robot in steep slope environments. Also, data sets created during different seasons should allow for automated documentation of the vineyard conditions. A further topic is the automated switching between trellis lines. The robot relies on the steel cable to secure itself from falling. However, the operation space becomes limited to a single row. Therefore, an autonomous carrier as presented in [Wo20] should move the robot automatically to the next row or vineyard, respectively.

Bibliography

[Ad17] Adamides, George; Katsanos, Christos; Parmet, Yisrael; Christou, Georgios; Xenos, Michalis; Hadzilacos, Thanasis; Edan, Yael: HRI usability evaluation of interaction modes for a teleoperated agricultural robotic sprayer. Applied Ergonomics, 62:237–246, 2017.

[Ar98] Arkin, R.: Behavior-Based Robotics. MIT Press, Cambridge, MA, USA, 1998.

[Ba21] Barker, Michael: An agricultural robot Slopehelper challenges steep vineyards and orchards. https://www.agricultural-robotics.com/news/an-agricultural-robot-slopehelper-challenges-steep-vineyards-and-orchards, May 2021.

[Br86] Brooks, R. A.: A Robust Layered Control System For A Mobile Robot. IEEE Journal of Robotics and Automation, 1:14–23, March 1986.

[Br13] Broggi, Alberto; Cardarelli, Elena; Cattani, Stefano; Sabbatelli, Mario: Terrain mapping for off-road Autonomous Ground Vehicles using rational B-Spline surfaces and stereo vision. In: 2013 IEEE Intelligent Vehicles Symposium (IV). pp. 648–653, 2013.

[Br17] Brunner, David: Untersuchungen zum Lenkverhalten eines fahrerlosen Trägerfahrzeugs mit Stachelwalzenantrieb für Weinbausteillagen. Dissertation, Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt. Lehrstuhl für Agrarsystemtechnik, München, 2017.

[Ga19] Gao, Biao; Xu, Anran; Pan, Yancheng; Zhao, Xijun; Yao, Wen; Zhao, Huijing: Off-Road Drivable Area Extraction Using 3D LiDAR Data. In: 2019 IEEE Intelligent Vehicles Symposium (IV). pp. 1505–1511, 2019.

[Io20] Iovino, Matteo; Scukins, Edvards; Styrud, Jonathan; Ögren, Petter; Smith, Christian: A Survey of Behavior Trees in Robotics and AI, 2020.

[Is05] Isla, Damian: GDC 2005 Proceeding: Handling Complexity in the Halo 2 AI. https://www.gamasutra.com/view/feature/2250/gdc_2005_proceeding_handling_.php, March, 11 2005.

[Jo04] Jones, J. L.: Robot Programming: A practical guide to Behavior-Based Robotics. McGraw-Hill, 2004.

[KFP17] Kodrič, Martin; Flašker, Jože; Pehan, Stanislav: Efficiency Improvement of Agricultural Winch Machines. Strojniški vestnik - Journal of Mechanical Engineering, 63(3), 2017.

[KGB21] Kötting, Christian; Gassen, Eike; Berns, Karsten: A Robot Platform for Steep Slope Vineyards. In: Agriculture Digitalization and Organic Production (2021). 2021.

[Ma90] Matarić, Maja J.: A Distributed Model for Mobile Robot Environment-Learning and Navigation. Technical report, USA, 1990.

[Ro18] Roure, Ferran; Moreno, Germán; Soler, Marcel; Faconti, Davide; Serrano, Daniel; Astolfi, Pietro; Bardaro, Gianluca; Gabrielli, Alessandro; Bascetta, Luca; Matteucci, Matteo: GRAPE: Ground Robot for vineyArd Monitoring and ProtEction. In (Ollero, Anibal; Sanfeliu, Alberto; Montano, Luis; Lau, Nuno; Cardeira, Carlos, eds): ROBOT 2017: Third Iberian Robotics Conference. Springer International Publishing, Cham, pp. 249–260, 2018.

[RWB17] Ropertz, Thorsten; Wolf, Patrick; Berns, Karsten: Quality-Based Behavior-Based Control for Autonomous Robots in Rough Environments. In (Gusikhin, Oleg; Madani, Kurosh, eds): Proceedings of the 14th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2017). volume 1, SCITEPRESS – Science and Technology Publications, Lda, Madrid, Spain, pp. 513–524, July 26–28 2017.

[Sa18] Santos, Luís; dos Santos, Filipe Neves; Mendes, Jorge; Ferraz, Nuno; Lima, José; Morais, Raul; Costa, Pedro: Path Planning for Automatic Recharging System for Steep-Slope Vineyard Robots. In (Ollero, Anibal; Sanfeliu, Alberto; Montano, Luis; Lau, Nuno; Cardeira, Carlos, eds): ROBOT 2017: Third Iberian Robotics Conference. Springer International Publishing, Cham, pp. 261–272, 2018.

[SBK12] Schwarz, H.-P.; Brunner, D.; Keicher, R.: Geisi - neues Fahrzeug für den Steillagenweinbau. Die Winzer-Zeitschrift, 26(3):27, 2012.

[SML21] Strub, Larissa; Mueller Loose, Simone: Kostennachteil Handarbeit in Steillagen. Der deutsche Weinbau, 2021(13):14–18, 06 2021.

[SPB08] Schäfer, B. H.; Proetzsch, M.; Berns, K.: Action/Perception-Oriented Robot Software Design: An Application in Off-road Terrain. In: 10th International Conference on Control, Automation, Robotics and Vision. Hanoi, Vietnam, pp. 223–228, Dezember 2008.

[WB21] Wolf, Patrick; Berns, Karsten: Data-fusion for robust off-road perception considering data quality of uncertain sensors. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague, Czech Republic, pp. 6853–6860, September 27–October 1 2021.

[Wo18] Wolf, Patrick; Ropertz, Thorsten; Berns, Karsten; Thul, Martin; Wetzel, Peter; Vogt, Achim: Behavior-Based Control for Safe and Robust Navigation of an Unimog in Off-Road Environments. In (K. Berns et. al., ed.): Commercial Vehicle Technology 2018. Proceedings of the 5th Commercial Vehicle Technology Symposium – CVT 2018. Springer Vieweg, Kaiserslautern, Germany, pp. 63–76, March 13–15 2018.

[Wo20] Wolf, Patrick; Vierling, Axel; Ropertz, Thorsten; Berns, Karsten: Advanced scene aware navigation for the heavy duty off-road vehicle Unimog. IOP Conference Series: Materials Science and Engineering, 997:012093, Dec 2020.

[Wo21] Wolf, Patrick; Vierling, Axel; Husemann, Jörg; Berns, Karsten; Decker, Peter: Extending Skills of Autonomous Off-Road Robots on the Example of Behavior-Based Edge Compaction in a Road Construction Scenario. In (K. Berns et. al., ed.): Commercial Vehicle Technology 2020/2021. Proceedings of the 6th Commercial Vehicle Technology Symposium – CVT 2020/2021. Commercial Vehicle Alliance Kaiserslautern (CVA), Springer Vieweg, Kaiserslautern, Germany, 2021.

AgriRegio: Infrastruktur zur Förderung von digitaler Resilienz und Klimaresilienz im ländlichen Raum am Beispiel der Pilotregion Nahe-Donnersberg

Christian Reuter[1], Franz Kuntke[1], Matthias Trapp[2], Christian Wied[3], Gerwin Brill[4], Georg Müller[5], Enno Steinbrink[1], Jonas Franken[1], Daniel Eberz-Eder[6], Wolfgang Schneider[7]

Abstract: Die Digitalisierung schreitet auch in der Landwirtschaft immer weiter voran. Vermehrt werden in landwirtschaftlichen Betrieben sogenannte Smart Farming-Technologien eingesetzt, mit deren Hilfe verschiedenste Arbeitsabläufe automatisiert ablaufen, kontrolliert werden und mit anderen Betrieben ausgetauscht werden können. Durch die verfügbaren Daten und die Vernetzung mit anderen Betrieben, ergeben sich vielfältige neue Möglichkeiten in Bezug auf ressourcenschonendes, wirtschaftlicheres und kollaboratives Arbeiten. Problematiken ergeben sich mit Blick auf die Speicherung dieser sensiblen Betriebsdaten, vor allem, wenn hierfür nur einige wenige Anbieter zur Verfügung stehen. Das Forschungsprojekt „AgriRegio“ soll die digitalisierte Datenerfassung und -nutzung in landwirtschaftlichen Betrieben widerstandsfähiger machen und die sicherheitskritische Infrastruktur schützen. Sieben Projektpartner erproben dazu smarte Sensoren auf Basis standardisierter Open-Source-Technologien in der Landwirtschaft, bei denen die Betriebsdaten dezentral auf lokalen Servern gespeichert werden.

Keywords: Resilienz, Edge Computing, Landwirtschaft

1 Einleitung

Beim sogenannten Smart Farming werden verschiedenste Sensoren, Geräte und Systeme eines landwirtschaftlichen Betriebs miteinander vernetzt. Dabei werden Daten genutzt, die entweder von amtlichen Stellen abgerufen oder selbst erzeugt werden. Auf diese Weise können beispielsweise Bodeneigenschaften, Landschaftsstrukturen oder Verkehrswege in die Planung einbezogen werden. Smart Farming macht also ein ressourcenschonendes und wirtschaftlicheres Arbeiten möglich. Zudem können die Daten beim überbetrieblichen Lo-

[1] Technische Universität Darmstadt (TUDa), Wissenschaft und Technik für Frieden und Sicherheit (PEASEC), Pankratiusstr. 2, 64298 Darmstadt, <nachname>@peasec.tu-darmstadt.de, https://orcid.org/0000-0003-1920-038X, https://orcid.org/0000-0002-7656-5919
[2] RLP AgroScience, Breitenweg 71, 67435 Neustadt an der Weinstraße, matthias.trapp@agroscience.rlp.de
[3] IBM Deutschland GmbH, IBM-Allee 1, 71139 Ehningen, christian.wied@de.ibm.com
[4] expeer GmbH, Wachsbleiche 10-12, 53111 Bonn, brill@expeer.de
[5] Maschinen- und Betriebshilfsring Rheinhessen-Nahe-Donnersberg e.V., Otto-Lilienthal-Str. 4, 55232 Alzey, georg.mueller@mbr-alzey.de
[6] Land Rheinland-Pfalz, DLR Rheinhessen-Nahe-Hunsrück, Rüdesheimer Str. 60 - 68, 55545 Bad Kreuznach, daniel.eberz@dlr.rlp.de
[7] Maschinen- und Betriebshilfsring Rheinhessen-Nahe-Donnersberg e.V. (Mitglied), Otto-Lilienthal-Str. 4, 55232 Alzey, wolfgang.schneider@steinwein.de

doi:10.18420/inf2022_81

gistik- und Bewirtschaftungsmanagement neue Handlungsoptionen zum Boden-, Pflanzen- und Klimaschutz liefern. Dieses „datenlastige“ Arbeiten hat jedoch auch Nachteile, vor allem, wenn zur Datenspeicherung nur wenige oder ein einzelner Anbieter genutzt werden. Ein Ausfall oder gar Angriff auf die dabei eingesetzten digitalen Lösungen könnte dann im Extremfall zu Produktionsausfällen und Versorgungsengpässen führen. Hier setzt AgriRegio an.

Im Rahmen des Projekts „AgriRegio – Infrastruktur zur Förderung von digitaler Resilienz und Klimaresilienz im ländlichen Raum am Beispiel der Pilotregion Nahe-Donnersberg“ werden mit Hilfe standardisierter Open-Source-Technologien neue Anwendungen entwickelt, mit denen die beteiligten Landwirtschaftsbetriebe vor Ort ihre Daten erfassen und verarbeiten können, beispielsweise die Bodenfeuchtigkeit. Werden mehrere Sensoren eingesetzt, die sich miteinander verbinden, spricht man von Sensornetzen. Sensornetze können auch überbetrieblich im Verbund betrieben werden. Die Daten werden dabei per Funkverbindung direkt an die Betriebe übermittelt, und sind damit vom Internet unabhängig.

Gespeichert werden die Daten lokal auf einem kostengünstigen Miniserver, der „HofBox“, die im Rahmen des Vorgänger-Projekts „GeoBox-Infrastruktur“ entstanden ist und parallel weiterentwickelt wird. Hier befinden sich auch Betriebsdaten, Geodaten und Fachinformationen. Weil die Daten nicht auf einem zentralen Server gespeichert werden, sondern dezentral bei jedem landwirtschaftlichen Betrieb in der jeweiligen HofBox, entsteht eine widerstandsfähige digitale Infrastruktur, die zudem auch bei Internetausfällen funktionsfähig bleibt.

Ziel des Projekts AgriRegio ist es zu zeigen, dass sich auf regionaler Ebene ein flächendeckendes und vom Internet unabhängiges Sensornetz mithilfe von aktuellen, standardisierten Technologien des Edge Computings etablieren lässt. Das innovative Konzept für dezentrales und resilientes Edge Computing soll zur Praxisreife entwickelt und erprobt werden. Resilientes Edge Computing bedeutet dabei, dass digital erfasste Daten dezentral verarbeitet werden – ohne Anbindung an zentrale Server in Rechenzentren. Das Konzept wird im Rahmen einer öffentlich-privaten Partnerschaft unter Einbeziehung von regionalen Erzeugerorganisationen und Gebietskörperschaften in den nächsten zwei Jahren in der Pilotregion Nahe-Donnersberg in Rheinland-Pfalz getestet.

In diesem Artikel wird die Zielsetzung des Projekts, die bisherigen Arbeiten, der Forschungs- und Entwicklungsansatz sowie die Erfolgsaussichten besprochen.

2 Hintergrund

Nach der Verordnung zur Bestimmung kritischer Infrastrukturen gilt ein Landwirt als Betreiber kritischer Infrastruktur im Sektor Ernährung, sofern der Schwellenwert der Produktion von 434.500 t Speisen oder 350 Mio. Liter Getränke pro Jahr überschritten wird. Ein Angriff auf die in der Landwirtschaft weit verbreiteten IT-Systeme würde aber weitaus mehr Betriebe treffen, als die per Gesetz als KRITIS eingestuften. Es könnte somit zu

einer deutlich höheren Zahl an betroffenen Personen führen, als der vom BSI zu Grunde gelegte Schwellenwert von 500’000 versorgten Personen. Gleichzeitig gilt: Im Ernährungssicherstellungs- und -vorsorgegesetz (ESVG) ist die Deckung des lebensnotwendigen Nahrungsmittelbedarfs im Falle einer Versorgungskrise durch den Staat sicherzustellen.

Der Fokus der Landwirtschaft liegt insbesondere auf der präzisen und nachhaltigen Bewirtschaftung des Bodens. Dies soll künftig das *Smart Farming* durch Erhebung und Analyse von Prozess- und Sensordaten ermöglichen. Die derzeit auf dem Markt verfügbaren Dienstleistungen und Produkte sind dabei durch die Funktionsweise des Cloud Computing geprägt. Das bedeutet, dass Daten nicht mehr vor Ort gespeichert werden, sondern auf Servern in Rechenzentren ausgelagert werden. Bei den betrieblichen Daten handelt es sich um freiwillig bereitgestellte Betriebsgeheimnisse [LKSF21]. Das heißt, dass das Anwenden von Sanktionen schwierig ist, wenn ein Cloud-Anbieter diese Daten für einen nicht legitimen Zweck verwendet [LSKF22]. Zusätzlich ist es schwer nachzuvollziehen, wie die Daten von einem Cloud-Anbieter innerhalb seiner eigenen Rechnersysteme verwendet werden.

Ein weiteres Problem stellt die Ausfallsicherheit der Vernetzung dar. Da der Service vieler Anbieter zumeist wie eine zentrale Drehscheibe funktioniert, über die alle Aktionen innerhalb des landwirtschaftlichen Betriebes koordiniert werden, muss bei deren Ausfall im schlimmsten Falle die gesamte Geschäftstätigkeit stillgelegt werden. Nutzen also ausreichend viele große Betriebe den gleichen Anbieter, so kann es im Extremfall zu Produktionsausfällen beziehungsweise Versorgungsengpässen kommen. Auch absichtlich verursachte Ausfälle durch Cyberangriffe sind nicht auszuschließen (zum Beispiel Denial-of-Service-Angriffe [RSEB18]).

Eine mögliche Gegenmaßnahme, um die Gefahren des Cloud-Computing zu adressieren wäre hier, zumindest ein teilweise eigenes dezentrales Netzwerk zu errichten (zum Beispiel „Digitale HofBox“ [KRSE20]). Bei einem solchen „Offline-First“-System geht es darum, dass Programme grundsätzlich ohne Internetanbindung nutzbar sind. Sie können zusätzlich auch noch alle gewohnten Online-Fähigkeiten bieten, um so beispielsweise eine optionale Steuerung über das Smartphone zu ermöglichen [KLSF22]. Um eine möglichst resiliente Infrastruktur zu gewährleisten, ist ein internes Rechner-zu-Rechner System einer zentralisierten Cloud-Lösung in jedem Fall vorzuziehen.

Die Vernetzung und Digitalisierung in der Ernährungswirtschaft nimmt exponentiell zu und wird große Veränderungen bringen. Der kritischen Infrastruktur Landwirtschaft und der notwendigen kritischen Auseinandersetzung mit dem Sicherheitsaspekt der Technologie muss ausreichend Aufmerksamkeit eingeräumt werden. Es wäre daher von hoher Relevanz, die Infrastruktur für ein resilientes Smart Farming (RSF) zu erstellen. Ziel ist es daher, die Fortschritte der Digitalisierung in der Landwirtschaft zu nutzen, ohne die Ausfallsicherheit der landwirtschaftlichen Primärproduktion und damit die Lebensmittelversorgung der Verbraucher zu gefährden. Das Projekt AgriRegio möchte hierzu einen Beitrag leisten.

3 Zielsetzung des Projekts AgriRegio

Ziel des Projekts AgriRegio ist es, bei der Digitalisierung in der Landwirtschaft neue und innovative Konzepte eines dezentralen und resilienten Edge Computings (REC) [EKSR21] zur Praxisreife zu entwickeln und in der rheinland-pfälzischen Pilotregion Nahe-Donnersberg im Rahmen einer überbetrieblichen Vernetzung in öffentlich-privater Partnerschaft unter Einbeziehung von regionalen Erzeugerorganisationen und Gebietskörperschaften zu testen. Im Sinne der experimentellen Forschungsmethodik sollen die Projektpartner die finalen, konkreten Lösungen im Zuge des Projekts integrieren.

Das resiliente Edge Computing dient der Weiterentwicklung der GeoBox-Infrastruktur [KRSE20] zur dezentralen Datenhaltung und regionalen Vernetzung, die das BMEL mit Zustimmung der Agrarministerkonferenz seit 2018 über den Projektträger BLE und die Landwirtschaftliche Rentenbank fördert. Zum bisherigen Förderbezug, der sich primär auf die Sicherung der betrieblichen Datenhoheit und Wertschöpfung durch die Etablierung von offline-fähigen HofBoxen richtete, bringt AgriRegio mit der Anbindung von IoT-Sensornetzwerken mit innovativer Analyse der Sensordaten nahe am Punkt ihrer Erzeugung, sprich am Netzwerkrand oder der „Edge“, zusätzliche Optionen eines überbetrieblich verankerten Umwelt- und Klimaschutzes auf regionaler Ebene.

AgriRegio soll demonstrieren, dass sich IoT-Sensoren auf Basis aktueller und standardisierter Open Source-Technologien des Edge Computings an die HofBoxen der GeoBox-Infrastruktur andocken lassen und Landwirte damit eigene IoT-Sensornetze im überbetrieblichen Verbund betreiben können. Die Koordination entsprechender Erzeugernetzwerke mit sensorgestützten HofBoxen erfolgt in AgriRegio exemplarisch durch eine Allianz des Informationsdienstleisters RLP AgroScience mit dem regional zuständigen Maschinenring Alzey, der die gewonnenen Sensordaten in Form von High-Level-Services nutzen kann. Beispielsweise bieten Sensordaten zur Ableitung der Bodenfeuchte und Blattnässe den Landwirten nicht nur betriebsinterne Vorteile, sondern liefern auch beim überbetrieblichen Logistik- und Bewirtschaftungsmanagement des Maschinenrings per Dashboard eine neue Dimension an regionalen Handlungsoptionen zum Boden-, Pflanzen- und Klimaschutz sowie zur Verbesserung der Wirtschaftlichkeit eines überbetrieblichen Smart Farmings.

Ausgehend vom Hofgut Neumühle in Münchweiler, das als „Open Data Farm“ auf Basis der GeoBox-Infrastruktur im Zentrum des Experimentierfelds Südwest steht und mit dem BMEL-Projekt SOFI (Smart Soil Information for Farmers) zur Bodenfeuchteabschätzung bereits bundesweit auf die KlimAgrar-Projekte ausstrahlt, soll in AgriRegio gemeinsam mit weiteren Betrieben in der Pilotregion Nahe-Donnersberg ein regionales IoT-Sensornetzwerk per resilienter LoRaWAN-Kommunikation aufgebaut und getestet werden [KRLS22]. In AgriRegio kann der beteiligte Maschinenring als Selbsthilfeorganisation der Landwirte den Nachweis erbringen, dass sich mit dem resilienten Edge Computing auf regionaler Ebene ein flächendeckendes und vom Internet unabhängiges Sensornetz etablieren lässt. Entscheidend ist hierbei der in AgriRegio verfolgte Open Source-Ansatz, der ein Rollout der sensorgestützten HofBoxen nicht durch Schutzrechte beeinträchtigt.

Mit diesem Aufbau ausfallsicherer Erzeugernetzwerke und der Organisationskompetenz beim überbetrieblichen Logistik- und Bewirtschaftungsmanagement ist der Maschinenring prädestiniert, um in der kritischen Infrastruktur (KRITIS) „Ernährung“ den in Krisenfällen regional zuständigen Gebietskörperschaften als Verwaltungshelfer gemäß des Ernährungssicherstellungs- und -vorsorgegesetzes (ESVG) zuzuarbeiten. Um der wachsenden Vulnerabilität einer internetabhängig digitalisierten Landwirtschaft entgegenzuwirken, werden in AgriRegio erste Testläufe des resilienten Edge Computing der GeoBox-Infrastruktur bei simulierten Internetausfällen in der Pilotregion Nahe-Donnersberg durchgeführt.

Die Ausrichtung der Resilienzszenarien und darauf abgestimmte Testläufe werden in enger Zusammenarbeit mit Vertretern der assoziierten Kreisverwaltungen im Rahmen regelmäßiger Workshops festgelegt und gemäß den implementierten Technologien dynamisch weiterentwickelt und umgesetzt.

Auf Forschungsebene soll erstmalig in Zusammenarbeit mit den assoziierten Kreisverwaltungen evaluiert werden, ob regionale und resiliente IoT-Sensornetzwerke auf Basis der in AgriRegio fusionierten REC- und LoRaWAN-Technologien neben dem direkten Nutzen beim Smart Farming und dem Klimaschutz auch der digitalen Resilienz im Rahmen gesellschaftlich relevanter ESVG-Krisenpläne zur Ernährungssicherstellung dienen können. Eine derartige „Doppelnutzung” ist politisch willkommen und würde der flächendeckend operierenden Landwirtschaft einen neuen Stellenwert bei der im GeoBox-Projekt u.a. mit IBM initiierten öffentlich-privaten Partnerschaft zur Etablierung einer "digitalen Resilienz als Service“ (dRaaS) in der Kritischen Infrastruktur Ernährung verschaffen. Die in AgriRegio geplante experimentelle Operationalisierung einer resilienten Infrastruktur bei Internetausfällen ist offen für eine herstellerübergreifende Beteiligung der IT-Branche und insbesondere für Anbieter vulnerabler Cloudlösungen von zunehmender Relevanz. Es werden keine Schutzrechte durch die Arbeiten im Rahmen von AgriRegio angestrebt.

4 Bisherige Arbeiten

Koordiniert wird das Verbundprojekt vom Fachgebiet Wissenschaft und Technik für Frieden und Sicherheit (PEASEC) am Fachbereich Informatik der Technischen Universität Darmstadt. Weitere Partner sind IBM Deutschland GmbH, expeer GmbH, RLP AgroScience, der Maschinen- und Betriebshilfsring Rheinhessen-Nahe-Donnersberg e.V. sowie die Kreisverwaltungen Donnersbergkreis und Bad Kreuznach sowie das Dienstleistungszentren Ländlicher Raum (DLR) Rheinhessen-Nahe-Hunsrück. Das BMEL fördert AgriRegio für zwei Jahre aus Mitteln der Landwirtschaftlichen Rentenbank.

Der Lehrstuhl *Wissenschaft und Technik für Frieden und Sicherheit (PEASEC)* im Fachbereich Informatik der *Technischen Universität Darmstadt* verbindet Informatik mit Sicherheitsforschung. Im Fach Informatik gehört die TU europaweit zu den führenden Universitäten – gemessen an der Forschungsleistung. PEASEC forscht zur resilienten Digitalisierung in der Landwirtschaft und beschäftigt sich mit der Nutzbarkeit von Sicherheits-

und Privatheitskonzepten [LiVR19, LKSR19] zum Schutz vor gezielten Angriffen und unbeabsichtigten Fehlfunktionen. PEASEC ist beteiligt an dem BMBF-Projekt *HyServ* (BMBF; Hybride Dienstleistungen) [BBSL21] sowie den BMEL-Projekten *GeoBox-I* (Dezentrale Datenhaltung und Regionale Vernetzung) und *GeoBox-II* (Standardisierter Datenaustausch mit amtlichen Informationsportalen) sowie Koordinator des Forschungsbereichs *Secure Urban Infrastructures* im Nationalen Forschungszentrum für angewandte Cybersicherheit ATHENE.

IBM gehört mit einem Umsatz von 73.620 Millionen US-Dollar im Jahr 2020 zu den weltweit größten Anbietern im Bereich Informationstechnologie (Hardware, Software und Services) und B2B-Lösungen. *IBM Deutschland* ist an rund 15 Standorten in ganz Deutschland vertreten und sichert damit nicht nur in allen wirtschaftlichen Metropolregionen, sondern auch in den Mittel- und Regionalzentren kurze Wege und beste Erreichbarkeit und Kundennähe. Künstliche Intelligenz/Business Analytics, Cloud Computing, Security, das Nutzen der Möglichkeiten im Internet der Dinge (IoT) und Blockchain-Technologien sind die strategischen Felder und bilden für IBM die Basis ihres stetig erweiterten Lösungsportfolios. IBM Edge Computing wurde entwickelt, zur Skalierung und zum Betrieb von Edge-Lösungen in jeder Umgebung mit autonomem Management. Eine wesentliche Zielsetzung ist autonome Ausführung, auch im Offlinebetrieb zur Reduzierung von Unterbrechungen und Kosten.

Die *expeer GmbH* ist ein gründergeführtes IT-Technologie- und Beratungsunternehmen aus Bonn, das überwiegend im Bereich der digitalen Transformation tätig ist. Seit 2002 setzt expeer hochkomplexe Projekte um. So sind die eingesetzten Prozesse seit 2017 nach ISO 9001:2015 zertifiziert. Seit 2019 ist expeer ein autorisiertes Beratungsunternehmen im vom Bundesministerium für Wirtschaft und Energie getragenen Programm go-digital. FuE-Tätigkeiten, u.a. für die Deutsche Luft- und Raumfahrtgesellschaft sowie die Bundesanstalt für Landwirtschaft und Ernährung, zeigen auf, mit expeer einen zukunftsorientierten Partner an der Seite zu haben. Die von expeer u.a. in den Forschungsprojekten TruNet sowie GeoBox-I/II mitentwickelten GeoBox-Technologien verknüpfen die Produktionsdaten landwirtschaftlicher Betriebe mit öffentlich bereitgestellten Informationen und sind somit zugleich redundanter lokaler (Zwischen-) Speicher und Drehscheibe im Prozess der landwirtschaftlichen Digitalisierung und ländlichen Vernetzung. Auf Grundlage der von expeer eingebrachten Basistechnologien jim!, jim!goesMobile und jim!trustNet-Messenger realisieren wir einen dezentralen inhalts-agnostischen Enterprise-Service-Bus, der mit dem GeoBox-Messenger als Bedienoberfläche prinzipiell beliebige Anwendungsfälle und Workflows realisieren kann. P2P-basiertes WebRTC-Messaging in Kombination mit den dezentralen HofBoxen bietet zudem die Möglichkeit zur Untersuchung einer resilienten Inselnetz-Kommunikation.

Der *Maschinen- und Betriebshilfsring Rheinhessen-Nahe und Donnersberg e.V.* (MR) mit Sitz in Alzey übernimmt als Selbsthilfeorganisation der Landwirte auf regionaler und überbetrieblich ausgerichteter Ebene eine wichtige Rolle bei der digitalen Transformation der Landwirtschaft. Mit rund 2600 Mitgliedsbetrieben ist es eine der größten Maschinen-

ringeinheiten auf Bundesebene. Das Arbeitsgebiet umfasst 2.900 km^2 mit 150.600 ha landwirtschaftlicher Nutzfläche in 337 Gemeinden mit 840.000 Einwohnern. Neben der Organisation und Abrechnung überbetrieblicher Arbeiten entwickelt sich der Ausbau eines regional vernetzten Datenmanagements zwischen den Mitgliedsbetrieben zur Zukunftsaufgabe. Eine besondere Herausforderung stellt hierbei der arbeitsteilige Einsatz von informationsgestützter Landtechnik (überbetriebliches Smart Farming) dar. Am Beispiel des MR-Mitgliedbetriebs Hofgut Neumühle (u.a. „Open Data Farm" im Experimentierfeld Südwest) entwickelte der Maschinenring gemeinsam mit PEASEC im BMBF-Projekt HyServ [BRSL19] eine blockchainbasierte Auftragsabwicklung für hybride, d.h. informationsgestützte Maschinendienstleistungen. Satzungsgemäße Aufgabe der Maschinenringe in RLP ist auch die Unterstützung der Gesellschaft in Krisenfällen. Entsprechend groß ist das Interesse an einer resilienten Form der Digitalisierung in der landwirtschaftlichen Primärproduktion, um ggf. als „Verwaltungshelfer" (n. ESVG, §3 Abs. 4) die für die regionale Ernährungssicherstellung zuständigen Landkreise in Krisenfällen mit möglichst ausfallsicheren Digitalisierungslösungen unterstützen zu können.

Gegenstand der *RLP AgroScience GmbH* sind grundlagen- und anwendungsorientierte Forschungs- und Entwicklungsarbeiten auf den Gebieten Biotechnologie und Agrarökologie im Zusammenhang mit Pflanzenproduktion. Die gemeinnützige Gesellschaft verfolgt das Ziel, diese Wirtschaftsweisen ökonomisch und ökologisch zu optimieren und das erarbeitete Wissen Unternehmen, Forschungseinrichtungen und interessierten Personen zur Verfügung zu stellen. Die Gesellschaft soll insbesondere die Wettbewerbsfähigkeit von Sonderkulturbetrieben (Weinbau, Obstbau, Gartenbau und sonstige) sowie von verarbeitenden Betrieben unterstützen und den Umwelt- und Naturschutz, vor allem in landbaulich intensiv genutzten Regionen, durch ihre Forschungstätigkeit fördern. Das Kernthema Anwendungen der Digitalisierung setzt Methoden der digitalen Landschaftsanalyse ein, um den landscape data space als raumzeitlichen Kontext weiter zu etablieren und als ein zentrales Element in die Geoboxinfrastruktur des Landes zu integrieren. Die RLP AgroScience ist Konsortialführer des Forschungsvorhaben SOFI sowie Forschungspartner im Verbundprojekt Experimentierfelder Südwest (beides gefördert durch das BMEL).

5 Forschungs- und Entwicklungsansatz

In AgriRegio gilt es den Nachweis zu führen, dass ein regional und überbetrieblich ausgerichteter Digitalisierungsansatz in der Landwirtschaft sofort im Tagesgeschäft seinen Nutzen entfalten kann und sich doch im Krisenfall, d.h. bei einem länger andauernden Internetausfall, in einem Notmodus weiterverwenden lässt. Diesen Anspruch kann die Geo-Box-Infrastruktur durch die in AgriRegio geplanten Entwicklungsschritte zur Integration des resilienten und sensorgestützten Edge Computings inzwischen immer besser erfüllen. Zu betonen ist hierbei, dass die herstellerunabhängige regionale Vernetzung von HofBoxen mit standardisierter Datenhaltung per 0/5-G-Kommunikation (d.h. inkl. LoRaWAN) nicht im Wettbewerb zu den proprietären und vielfach vulnerablen Cloudlösungen

der Wirtschaft steht. Vielmehr erlaubt die Etablierung der resilienten und von der Politik geforderten bzw. geförderten resilienten GeoBox-Infrastruktur als inselnetzfähiges „Fallback-System“ eine weitere Verwendung cloudbasierter Farmmanagementsysteme in der KRITIS Landwirtschaft und Ernährung [ReSE19].

Da sich in der Landwirtschaft bisher noch keine allgemeingültigen Datenaustauschstandards herauskristallisiert haben, arbeitet das vom BMEL geförderte GeoBox-Vorhaben an einer schema-agnostischen Datenhaltung in der HofBox, um Branchenlösungen auf Basis semantischer Technologien flexibel anbinden zu können. In diese Entwicklung sind die AgriRegio-Partner TUDa und das Softwarehaus expeer aktiv eingebunden, so dass die Berücksichtigung von aktuellen Schnittstellenentwicklungen sichergestellt ist. In dem für AgriRegio besonders relevanten Bereich der Geo- und Sensordaten wird die in der EU (INSPIRE) kürzlich beschlossene Umstellung auf Anwendungsprogrammierschnittstellen (z.B. OpenAPI, OGC API Features, SensorThings API) direkt umgesetzt.

Eine besondere Herausforderung beim dezentralen Edge Computing ist die Auslieferung und der fortlaufende Support der Edge Devices (HofBoxen). Hierzu kann AgriRegio erstmals auf das Open Source-Framework Open Horizon der Linux Foundation zurückgreifen. IBM fördert diese Open Source-Entwicklung für das Edge Computing und stellt in AgriRegio die Infrastruktur zur automatisierten Installation der HofBoxen (SDO-Standard: Secure Device Onboarding) und zur Integration und Fernwartung vielfältiger Softwarecontainer zur sicheren Offline-Nutzung in den Betrieben bereit. Im Projektverlauf von AgriRegio entwickelt und installiert IBM exemplarische Auslieferungskanäle für offline-fähige Softwareapplikationen, die im Sinne der digitalen Resilienz als Service (dRaaS) branchenweit auch von Drittanbietern genutzt werden können. Der Support von HofBoxen in der GeoBox-Infrastruktur wird über transportable und automatisierte Edge-Cloud-Rechenzentren abgesichert, die dezentral auf die für die Ernährungssicherstellung zuständigen Gebietskörperschaften (i.d.R. Landkreise) verteilt sind. Als Hardware für die betriebliche HofBox dienen aus Kostengründen primär Einplatinencomputer (z.B. Raspberry Pi), aber auch leistungsfähige Server oder Windows PCs können per Open Horizon als HofBox genutzt werden.

Der Maschinenring entwickelt und praktiziert ein entsprechendes Coaching für die Landwirte als HofBox-Nutzer und unterstützt diese beim Aufbau des regionalen Sensornetzwerks, welches eine internetunabhängige LoRaWAN-Installation zur Datenübertragung nutzt, vergleichbar zu [KSLR21]. Aufgrund der besonderen Umwelt- und Klimarelevanz der Bodenfeuchte beim (überbetrieblichen) Maschineneinsatz fokussiert sich der exemplarische Sensoreinsatz im AgriRegio-Vorhaben zunächst auf die Parameter Bodenfeuchte zur Ableitung der Befahrbarkeit von Böden sowie Lufttemperatur und Luftfeuchtigkeit zur Steuerung von überbetrieblichen Pflanzenschutzmaßnahmen.

Das Ausrollen der HofBoxen wird durch Evaluationsstudien der TUDa begleitet. Dabei werden mögliche Akzeptanzhemmnisse und Probleme seitens der Nutzerführung sowohl bei Inbetriebnahme, als auch fortwährender Nutzung der HofBoxen ermittelt. Im Rahmen

der Verbesserung der HofBox-Software wird die Forschung an benutzbaren privatheitsfördernden Technologien [LKSF21] fortgeführt.

Zur Datenauswertung der betrieblichen und regionalen Sensordatenströme entwickelt die RLP AgroScience ausgehend vom BMEL-Forschungsvorhaben SOFI das Bodenfeuchtesimulationsmodel SIMPEL weiter, um generische und mögliche Niederschlags- und Temperaturzeitreihen zu kreieren und deren Auswirkung auf den Einzelschlag zu rechnen [TrTK13].

Als weitere Anwendung werden aufbauend auf Sensornetzen zur kontinuierlichen Erfassung von Lufttemperatur und Luftfeuchte Korrelationen zur aktuellen Blattnässe erarbeitet, um damit wiederum Pflanzenschutzmaßnahmen überbetrieblich optimieren zu können. Dazu werden geostatistische Methoden und machine learning-Algorithmen genutzt, die Sensordaten, Vor-Ort-Messungen und die stationsgebundenen Daten der Agrarmeteorologie miteinander in Verbindung setzen. Als Ergebnis sollten dann wiederum einfache Ableitungen und Modelle stehen, die als entsprechende Softwarecontainer aufgrund des resilienten Edge Computings auf der HofBox laufen können und zusätzlich lokale Bewirtschaftungsdaten in die Auswertung einbeziehen. Dabei behalten Landwirte grundsätzlich die Datenhoheit. Sie bekommen jedoch die Option, ihre standort- und bewirtschaftungsspezifisch verfeinerten Daten für eine überbetriebliche Auswertung freizugeben und damit den Aufbau regionaler High-Level-Services zu unterstützen. Entsprechende Ergebnisse werden als thematische Karten einerseits betriebsbezogen an die HofBoxen zurückgeliefert und andererseits regionalisiert über das Portal des GeoBox-Viewers veröffentlicht, das vom ebenfalls in der Pilotregion angesiedelten Dienstleistungszentrum Ländlicher Raum in Bad Kreuznach bereits länderübergreifend betrieben wird.

Im Fokus von AgriRegio steht allerdings die Einbindung der sensorgestützten Bodenfeuchte-Services in zu entwickelnde Dashboards für den Maschinenring. Ziel ist ein überbetriebliches und umweltsensitives Logistik- und Bewirtschaftungsmanagement, das RLP AgroScience gemeinsam mit IBM entwickelt. Aus Gründen des Umwelt- und Klimaschutzes ist diese regionale Steuerungsebene von besonderer Relevanz, da der Maschinenring beispielsweise lokale Niederschlagsereignisse beim überbetrieblichen Maschineneinsatz eher großräumig umgehen kann als ein standortgebundener Einzelbetrieb.

Zur hierfür erforderlichen Kommunikation von georeferenzierten Auftrags- und Bewirtschaftungsdaten zwischen Landwirten, Lohnunternehmern und Maschinenring entwickelt expeer für den bundesweit verfügbaren GeoBox-Messenger eine Teil-App, d.h. einen speziell an das Maschinenring-Framework angepassten Dokumentenchat in Form eines messengerbasierten Workflows. Dabei stellt der GeoBox-Messenger als prinzipiell anwendungsfallunabhängiger Datenaustauschkanal eine optimale Plattform auch für die regionale Vernetzung dar; er kann hier die Rolle eines Enterprise Service Bus (ESB) einnehmen. Zudem kann der auf mobilen Endgeräten installierte GeoBox-Messenger durch seine inhärenten ESB- und Schnittstellenfähigkeiten beispielsweise auf die Datenhaltung in der HofBox, aber auch auf lokale IoT-Sensornetze effizient zugreifen. Der GeoBox-Messen-

ger ist damit in der Lage, prinzipiell alle in der Landwirtschaft tätigen Personen auf individuellen Wegen und leichtgewichtig in das resiliente digitale Ökosystem von AgriRegio einzubinden.

AgriRegio liefert mit der GeoBox-Infrastruktur auf Basis des resilienten Edge Computings gewichtige Alleinstellungsmerkmale für eine regional ausgerichtete Digitalisierung in der Landwirtschaft, sowohl für das Tagesgeschäft als auch für Krisenfälle: 1.) Softwarecontainer für die betriebliche HofBox zur Offline-Datenhaltung sowie für angeschlossene IoT-Sensornetzwerke lassen sich automatisiert ausliefern und aktualisieren; 2.) Sensornetze können in der Landwirtschaft für klimarelevante High-Level-Services flächendeckend aufgebaut werden und bieten mit LoRaWAN den Einstieg in eine internetunabhängige und resiliente Kommunikation [KSLR21]; 3.) Messengerfunktionen dienen einer effizienten und herstellerunabhängigen Anbindung aller regional relevanten Akteure landwirtschaftlicher Produktions- und Lieferketten an übergeordnete Organisationseinheiten.

AgriRegio untersucht, wie die Maschinenringarbeit von vorgenannten Funktionen profitiert und Umwelt- und Klimaziele mit modernster Technologie angegangen werden können. Politisch und gesellschaftlich relevant ist zusätzlich die Frage, inwieweit Grundfunktionen der GeoBox-Infrastruktur bei länger andauernden Internetausfällen dazu beitragen können, essentielle Prozesse der landwirtschaftlichen Primärproduktion im Digitalzeitalter aufrecht zu erhalten. Durch Simulation von Internetausfällen prüft der Lehrstuhl Wissenschaft und Technik für Frieden und Sicherheit (PEASEC) der TUDa gemeinsam mit dem Maschinenring, inwieweit die regionale Operationalisierung des dezentralen und resilienten Edge Computings den Maschinenring dazu befähigt, als ESVG-Verwaltungshelfer die für die Ernährungssicherstellung zuständigen Landkreise im Krisenfall zu unterstützen.

Die prinzipielle Aufteilung der Aufgaben ist dabei wie folgt: Die TUDa übernimmt die wissenschaftliche Projektkoordination und ist direkter Absprache mit dem MR Alzey für die Evaluationen der Entwicklungen verantwortlich. Die Entwicklung neuer Software-Artefakte bzw. Konfiguration und Einbettung bestehender Software wird durch expeer (GeoBox-Applikationen) und IBM (Infrastruktur) übernommen. Domänenspezifische Anwendungen werden durch RLP AgroScience weiterentwickelt.

6 Erfolgsaussichten und Ausblick

Die in AgriRegio geplante experimentelle Zusammenführung bisheriger Entwicklungen der GeoBox-Infrastruktur mit innovativen Open Source-Lösungen zur Verbreitung des dezentralen und resilienten Edge Computings dient dem modellhaften Aufbau einer Resilienzregion, die als Referenz und Leuchtturm mit zentralem Support von ausfallsicher vernetzten HofBoxen und IoT-Sensoren sowohl die Klimaresilienz als auch die digitale Resilienz in Krisenfällen fördert und damit im Kontext von gesellschaftlich relevanten Entscheidungen zur Digitalisierung in der KRITIS Ernährung und Landwirtschaft wichtige Grundlagen liefern kann. Die Durchführung und Evaluierung simulierter Ausfallszenarien

wird dabei zugleich auch den wissenschaftlichen Diskurs voranbringen. Das Vorhaben schließt aus wissenschaftlicher Sicht nahtlos an frühere Vorarbeiten an und überführt innovative Konzepte in praktisch demonstrierte Machbarkeit.

AgriRegio zielt auf eine öffentlich-private Partnerschaft mit landwirtschaftlichen Erzeugerorganisationen einerseits, und den für die Ernährungssicherstellung (nach ESVG) zuständigen regionalen Gebietskörperschaften andererseits. Der an AgriRegio beteiligte Maschinenring hat ein primäres Interesse an Managementtools zur effizienten Operationalisierung eines herstellerunabhängigen und überbetrieblichen Smart Farmings. Diese in AgriRegio konzipierte Digitalisierungsstufe muss resilient und krisentauglich umgesetzt werden, da Maschinenringe in Rheinland-Pfalz satzungsbedingt den Staat in Krisenfällen unterstützen müssen. Entsprechend sollen die betrieblichen Komponenten der GeoBox-Infrastruktur auf Basis eines resilienten Edge Computings der Branche grundsätzlich als Open Source-Software zur Verfügung gestellt werden und auch der regionalen privatwirtschaftlichen Datenkommunikation dienen. Eine privatwirtschaftliche Nutzung des modular per Softwarecontainer erweiterbaren regionalen AgriRegio-Ökosystems durch Dritte ist erwünscht und kann durch die IT-Projektpartner fachlich und technisch unterstützt werden. Schon zur Laufzeit des Projektes soll der Praxistransfer der Ergebnisse bei Kreisverwaltungen, Landwirten, Maschinenringen, Lohnunternehmern und weiteren Dienstleistern getestet und evaluiert werden.

Danksagung

Diese Arbeit wurde aus Mitteln des Zweckvermögens des Bundes bei der Landwirtschaftlichen Rentenbank im Rahmen des Projekts AgriRegio gefördert.

Literaturverzeichnis

[BBSL21] Bernardi, Ansgar ; Becker, Sandra ; Struve, Carsten ; Linsner, Sebastian ; Reuter, Christian ; Müller, Georg: Erfolgsorientierte Dienstleistung: Neue Perspektiven für die landwirtschaftliche Arbeitsteilung durch hybride Dienstleistungen. In: *41. GIL-Jahrestagung: Informatik in der Land-, Forst- und Ernährungswirtschaft* (2021), S. 37

[BRSL19] Bernardi, Ansgar ; Reuter, Christian ; Schneider, Wolfgang ; Linsner, Sebastian ; Kaufhold, Marc-André: Hybride Dienstleistungen in digitalisierten Kooperationen in der Landwirtschaft. In: Meyer-Aurich, A. (Hrsg.): *38. GIL-Jahrestagung: Digitalisierung in kleinstrukturierten Regionen, Lecture Notes in Informatics (LNI)*. Vienna, Austria : Gesellschaft für Informatik, 2019

[EKSR21] Eberz-Eder, Daniel ; Kuntke, Franz ; Schneider, Wolfgang ; Reuter, Christian: Technologische Umsetzung des Resilient Smart Farming (RSF) durch den Einsatz von Edge-Computing. In: *41. GIL-Jahrestagung: Informatik in der Land-, Forst- und Ernährungswirtschaft* (2021), S. 79–84

[KLSF22] Kuntke, Franz ; Linsner, Sebastian ; Steinbrink, Enno ; Franken, Jonas ; Reuter,

CHRISTIAN: Resilience in Agriculture: Communication and Energy Infrastructure Dependencies of German Farmers. In: *International Journal of Disaster Risk Science (IJDRS)* (2022)

[KRLS22] KUNTKE, FRANZ ; ROMANENKO, VLADIMIR ; LINSNER, SEBASTIAN ; STEINBRINK, ENNO ; REUTER, CHRISTIAN: LoRaWAN Security Issues and Mitigation Options by the Example of Agricultural IoT Scenarios. In: *Transactions on Emerging Telecommunications Technologies (ETT)* (2022)

[KRSE20] KUNTKE, FRANZ ; REUTER, CHRISTIAN ; SCHNEIDER, WOLFGANG ; EBERZ, DANIEL ; BERNARDI, ANSGAR: Die GeoBox-Vision: Resiliente Interaktion und Kooperation in der Landwirtschaft durch dezentrale Systeme. In: *Mensch und Computer 2020 - Workshopband* (2020), S. 1–6

[KSLR21] KUNTKE, FRANZ ; SINN, MARCEL ; LINSNER, SEBASTIAN ; REUTER, CHRISTIAN: Low Power Wide Area Networks (LPWAN) für krisentaugliche Datenübertragung in landwirtschaftlichen Betrieben. In: *41. GIL-Jahrestagung: Informatik in der Land-, Forst- und Ernährungswirtschaft* (2021), S. 193–198

[LiVR19] LINSNER, SEBASTIAN ; VARMA, RASHMI ; REUTER, CHRISTIAN: Vulnerability Assessment in the Smart Farming Infrastructure through Cyberattacks. In: MEYER-AURICH, A. (Hrsg.): *38. GIL-Jahrestagung: Digitalisierung in kleinstrukturierten Regionen, Lecture Notes in Informatics (LNI)*. Wien, Austria : Gesellschaft für Informatik, 2019

[LKSF21] LINSNER, SEBASTIAN ; KUNTKE, FRANZ ; STEINBRINK, ENNO ; FRANKEN, JONAS ; REUTER, CHRISTIAN: The Role of Privacy in Digitalization – Analysing the German Farmers' Perspective. In: *Proceedings on Privacy Enhancing Technologies (PoPETs)* Bd. 2021 (2021), Nr. 3

[LKSR19] LINSNER, SEBASTIAN ; KUNTKE, FRANZ ; SCHMIDBAUER-WOLF, GINA MARIA ; REUTER, CHRISTIAN: Blockchain in agriculture 4.0 - An empirical study on farmers expectations towards distributed services based on distributed ledger technology. In: *ACM International Conference Proceeding Series* Bd. 2019 (2019), S. 103–113 — ISBN 9781450371988

[LSKF22] LINSNER, SEBASTIAN ; STEINBRINK, ENNO ; KUNTKE, FRANZ ; FRANKEN, JONAS ; REUTER, CHRISTIAN: Transparency and Accountability in Privacy Oriented Data Disclosure Scenarios in Agriculture. In: *Behaviour \& Information Technology (BIT)* (2022)

[ReSE19] REUTER, CHRISTIAN ; SCHNEIDER, WOLFGANG ; EBERZ, DANIEL: Resilient Smart Farming (RSF) – Nutzung digitaler Technologien in krisensicherer Infrastruktur. In: VIENNA, A. (Hrsg.): *38. GIL-Jahrestagung: Digitalisierung in kleinstrukturierten Regionen, Lecture Notes in Informatics (LNI)*. Vienna, Austria : Gesellschaft für Informatik, 2019

[RSEB18] REUTER, CHRISTIAN ; SCHNEIDER, WOLFGANG ; EBERZ, DANIEL ; BAYER, MARKUS ; HARTUNG, DANIEL ; KAYGUSUZ, CEMAL: Resiliente Digitalisierung der kritischen Infrastruktur Landwirtschaft - mobil, dezentral, ausfallsicher. In: DACHSELT, R. ; WEBER, G. (Hrsg.): *Mensch und Computer 2018: Workshopband*. Dresden, Germany : Gesellschaft für Informatik e.V., 2018, S. 623–632

[TrTK13] TRAPP, MATTHIAS ; TINTRUP GEN. SUNTRUP, GREGOR ; KOTREMBA, CHRISTIAN: Auswirkungen des Klimawandels auf die Landwirtschaft und den Weinbau in Rheinland-Pfalz. In: *Schlussberichte des Landesprojekts Klima- und Landschaftswandel in Rheinland-Pfalz (KlimLandRP)* Bd. Teil 3 (2013), S. 170

Aufbau und Umsetzung eines informellen Netzwerkes zur Förderung der Digitalisierung in der Landwirtschaft in Rheinland-Pfalz zur Implementierung einer Wissenstransfer-Plattform

Lea Wintz [1], Paul Strerath[2]

Abstract: Precision Farming, Smart Farming bis hin zum Digital Farming: Die intensive Verwendung elektronischer Komponenten in landwirtschaftlichen Betrieben macht es heute möglich, vielfältige Daten aus unterschiedlichen Produktionsbereichen zu erfassen und verfügbar zu machen. Die Datenanalyse bietet das Potential, Betriebsmittel einzusparen, Dokumentationen zu erleichtern, die Entscheidungsfindung zu unterstützen und die Betriebsführung zu verbessern. Die Technische Hochschule Bingen untersucht im Rahmen des Verbundprojektes Experimentierfeld Südwest des Bundeslandwirtschaftsministeriums in Rheinland-Pfalz die Bedarfe der landwirtschaftlichen Betriebe in der Digitalisierung. Das Ziel ist, mittels Coaching und Schulung die Digitalisierung in landwirtschaftlichen Betrieben weiterzuentwickeln. Über ein herstellerunabhängiges, zentral gesteuertes informelles Netzwerk der Hochschule für angewandte Wissenschaft, wird der kritische Dialog zwischen Anwendern wie landwirtschaftlichen Betriebsleitern und Lohnunternehmern, Offizialberatung, Entwicklern und Herstellern von digitalen und KI gestützten Lösungen gefördert. Der individuelle Coaching-Bedarf einzelner Praktiker wird in Gesprächsrunden erkannt. Die als Ergebnis der Gesprächsrunden geschaffenen Wissenstransfer-Plattformen FarmWissen und OpenDataFarm werden als Kanäle des Coachings genutzt.

Keywords: Pilotbetriebe, Netzwerk, Digitalisierung, Landwirtschaft, Experimentierfeld Südwest

1 Einleitung

Die Digitalisierung der Landwirtschaft bietet Potenziale und Herausforderungen zur Steigerung der Effizienz der Betriebsmittel des Agrarsektors. Eine Optimierung der Nachhaltigkeit in den Bereichen Ressourcen- und Umweltschutz, sowie die Implementierung moderner Dokumentationsmethoden für eine transparentere Rückverfolgbarkeit von Lebensmitteln sind gesellschaftlich gewollt. Die aufgezählten Punkte führen zu einer Fokussierung auf eine Förderung der Digitalisierung in der

[1] Technische Hochschule Bingen, Fachbereich 1, Agrarwirtschaft, Berlinstraße 109, 55411 Bingen am Rhein, l.wintz@th-bingen.de
[2] Technische Hochschule Bingen, Fachbereich 1, Agrarwirtschaft, Berlinstraße 109, 55411 Bingen am Rhein, p.strerath@th-bingen.de

doi:10.18420/inf2022_82

Landwirtschaft. Diesem Ansinnen trägt das vom Bundesministerium für Ernährung und Landwirtschaft (BMEL) fachbereichsübergreifend geförderte Verbundprojekt „Förderung des branchenübergreifenden und überbetrieblichen Datenmanagements zur Unterstützung landwirtschaftlicher Wertschöpfungssysteme“ Rechnung. Wesentliches Ziel des Projekts ist eine Qualifizierung des landwirtschaftlichen Berufsstandes um digitale Technologien effizient einführen und nutzen zu können. Für die unterschiedlichen landwirtschaftlichen Anwendungsfälle werden Coaching Angebote erarbeitet und durchgeführt.

Vom Bundesministerium für Ernährung und Landwirtschaft (BMEL) werden insgesamt vierzehn Experimentierfelder im Bereich der Digitalisierung in der Landwirtschaft gefördert [Di21]. Das Projekt „Experimentierfeld Südwest“ (EF SW, https://ef-sw.de/) in Rheinland-Pfalz leistet einen Beitrag zur Förderung des branchenübergreifenden und überbetrieblichen Datenmanagements für die Unterstützung landwirtschaftlicher Betriebe. Charakteristisch für Rheinland-Pfalz ist eine kleinstrukturierte Agrarlandschaft mit vielfältigen Produktionsbereichen. Die kleine Agrarstruktur erschwert die Betriebsleitung vor dem Hintergrund des zunehmenden Strukturwandels. Zur Steigerung der Fachkompetenzen und der Akzeptanz digitaler Technologien seitens der Landwirte, soll der Umgang und der Nutzen bestehender digitaler Anwendungen gestärkt werden. Hierbei spielt der Wissenstransfer durch die Technische Hochschule Bingen (TH Bingen) eine wesentliche Rolle [Ri18] [Eb21].

In einem ersten Schritt wird der Stand der Digitalisierung der Landwirtschaft in Rheinland-Pfalz ermittelt. Dieses ist notwendig um weitere Schritte effektiv und effizient planen und umsetzen zu können. Die Digitalisierung in der Landwirtschaft legt ein hohes Tempo vor, so dass Datenerhebungen in der Regel sehr schnell veraltet sind. Der Digitalisierungsgrad der Landwirtschaft wird durch Einbeziehung der Praktiker (Landwirte, Obst- und Gemüseanbauer, Winzer, Lohnunternehmer, etc.) ermittelt. Hierbei steht weniger die statistische Belastbarkeit, als der „subjektive Überblick“ des Status-Quo im Fokus. Ein Einstieg in die Digitalisierung wird durch das Erstellen und Füllen einer Wissenstransfer–Plattform (FarmWissen) erleichtert.

Es werden Projektarbeiten, Bachelor- und Masterarbeiten zu spezifischen Fragen der Digitalisierung in der Landwirtschaft durchgeführt, mit der Zielsetzung die Ergebnisse in der regionalen Fachpresse zu veröffentlichen. Schulungsunterlagen werden sowohl im Curriculum des Studiums der Agrarwirtschaft wie auch in der beruflichen Bildung genutzt. Die Zielgruppe der Anwender wird weiterhin über die Fachpresse, Agrar- und Umwelttage der TH Bingen, regionale Agrarmessen und in- und ausländische Fachkongresse, über die Projektergebnisse informiert. Für die direkte Einbindung der landwirtschaftlichen Praxis, hat die TH Bingen ein informelles Netzwerk mit Pilotbetrieben verschiedener Produktionsbereiche aus Rheinland-Pfalz initiiert. Hierzu wurden Praktiker durch verschiedene Kanäle mehrfach angesprochen sich dem Netzwerk anzuschließen. Genutzt wurden die regionalen Leitmedien wie „Rheinische Bauernzeitung, Rheinland-Pfalz, Koblenz“, „DWZ – Die Winzer Zeitung, Rheinland-Pfalz, Koblenz“ und das „Wetterfax für Rheinhessen, DLR R-N-H, Bad Kreuznach“. Alleine die Rheinische Bauernzeitung hat eine kumulierte Nettoreichweite von 88% [St18]. Die Praktiker die sich nach dem Aufruf bei der TH Bingen gemeldet haben, bilden

das Netzwerk: Pilotbetriebe Digitalisierung. Das so geschaffene Netzwerk wird durch die wissenschaftlichen Mitarbeiter der TH Bingen betreut.
In diesem Beitrag werden die Ergebnisse bisheriger Befragungen durch die TH Bingen aufgeführt. Weiterhin werden Ergebnisse der durchgeführten Workshops im Netzwerk der Pilotbetriebe beschrieben. Hieraus wir der aktuelle Stand der Digitalisierung in der Landwirtschaft von Rheinland-Pfalz abgeleitet und wissenschaftliche Lösungsansätze zur Steigerung der Akzeptanz von digitalen Technologien werden aufgezeichnet. Ein Ausblick auf die weitere Arbeit in dem Projekt wird gegeben.

2 Stand der Forschung

2.1 Außenwirtschaft - Pflanzenbau

Um den Status-Quo der Digitalisierung in der Außenwirtschaft zu erfassen wurden zwanzig Betriebsleiter des Netzwerkes (Markfruchtbau 85 %, Weinbau 50 %, Grünland 15 %, Futterbau 10 %, Obstbau 10 %, Gemüsebau 5 %) befragt [Ku21]. Das Ziel war, die Meinungen und Erfahrungen zur Digitalisierung zu erfassen, in welche Technologien bereits investiert wurde und welche Probleme und Ängste mit dem Thema der Digitalisierung verbunden sind. Die Befragten waren durch folgende Eckdaten gekennzeichnet: Das mittlere Alter der Betriebsleiter betrug 39 Jahre, die durchschnittliche Betriebsgröße betrug 181 Hektar mit 2,6 Arbeitskräften. Einen Studienabschluss haben 40 % der Betriebsleiter, jeweils 20 % der Betriebsleiter sind staatlich geprüfte Landwirte, Techniker oder Meister; 50 % der Betriebe verfügen über eine Mobilfunknetzabdeckung von 3G. Die teilweise in Präsenz und digitale durchgeführte Befragung ergab eine positive Einstellung der Landwirte gegenüber der Digitalisierung mit dem Ziel, die Betriebe auf das digitale Zeitalter vorbereiten wollen. Vor allem Arbeitserleichterung, Einsparen von Kosten und das gleichzeitige Berücksichtigen von Umweltaspekten ist den Befragten wichtig (Tab.1).

Gründe	**Häufigkeit**	**in Prozent**
Arbeitserleichterung	19	95
Umweltaspekte berücksichtigen (Precision Farming)	18	90
Kosteneinsparung	16	80
Arbeitszeiteinsparung/Einsparen von Personal	14	70
Wirtschaftliche Vorteile	14	70
Landwirtschaft dem Verbraucher näher bringen	6	30
Mehrertrag im Anbau generieren	4	20

Tab. 1: Verteilung der Erwartungen der Betriebsleiter an die Digitalisierung [Ku21].

Welche positiven oder negativen Wirkungen, aufgrund der voranschreitenden Digitalisierung, sehen die Betriebsleiter voraus? Geringe Auswirkungen der zunehmenden Digitalisierung verspürten die Betriebsleiter bei dem möglichen Verlust von acker- und pflanzenbaulichen Expertisen. Dass die ständige Weiterentwicklung der digitalen Technik

zur Überforderung werden könnte, die Digitalisierung als Konfliktpotential zwischen den Generationen wahrgenommen und die digitale Technik eventuell überbetrieblich zur Kosteneinsparung genutzt werden wird, wurde als unproblematisch empfunden. Ausgeprägte Ängste haben die Betriebsleiter bei der Weitergabe von Betriebsdaten an Dritte. Sie fürchten einen Missbrauch der Daten und dadurch Nachteile für ihren Betrieb. Ebenfalls sind sie der Meinung, dass die Digitalisierung den Strukturwandel zu immer größeren Betrieben in der Landwirtschaft fördert, da sich die digitalen Technologien in Großbetrieben wirtschaftlicher einsetzen lassen (Tab. 2).

Probleme und Ängste	**Eher Ja %**	**Eher Nein %**
Verlust der ackerbaulichen Fähigkeiten	35	65
Durch Datenweitergabe an Dritte können Nachteile für die Betriebe entstehen	85	15
Die ständige Weiterentwicklung der digitalen Technik überfordert	30	70
Digitalisierung als Konfliktpotential zwischen Generationen	30	70
Digitale Technik lässt sich auf Großbetrieben wirtschaftlicher einsetzen -> fördert Strukturwandel	70	30
Digitale Technologien überbetrieblich nutzen oder in Maschinengemeinschaften eintreten -> Kosten sparen	45	55
Digitale Daten anonym zur Verfügung stellen	55	45
An einer Schulung zum Thema Digitalisierung teilnehmen und ggf. dafür zu bezahlen	80	20

Tab. 2: Mögliche Probleme und Ängste der Pilotbetriebe aufgrund der Digitalisierung [Ku21].

Der wesentliche begrenzende Faktor der voranschreitenden Digitalisierung ist nach wie vor eine nicht ausreichende Mobilfunknetzabdeckung in vielen landwirtschaftlichen Betrieben. Durch den freiwilligen Aufruf zu dieser Befragung, sind die Teilnehmenden eher als technikaffin zu bezeichnen. Die Pilotbetriebe repräsentieren ein vergleichsweise hohes Niveau an digitalen Technologien und oftmals besteht großes Interesse seitens der Betriebsleiter, die Digitalisierung in ihrem Betrieb zu erweitern. Durch die Gespräche mit den Betriebsleitern wurde weiterhin deutlich, dass mit zunehmender Digitalisierung die Kundenbindung zwischen Anbieter von digitalen Lösungen und landwirtschaftlichem Betrieb zunimmt [Ku21].

2.2 Innenwirtschaft - Milchviehhaltung

Die Anzahl an milchviehhaltenden Betrieben sinkt, bei gleichzeitig steigenden Herdengrößen [Bü19]. Das erforderliche Herdenmanagement kann durch eine verstärkte

Digitalisierung unterstützt werden [St20]. Um herauszufinden, auf welchem Digitalisierungsstand die Milchviehbetriebe in Rheinland-Pfalz und Luxemburg sind und welche Wünsche, Probleme und mögliche Verbesserungspotenziale bestehen, wurden 12 Landwirte aus Rheinland-Pfalz und drei Landwirte aus Luxemburg im Rahmen einer qualitativen Umfrage befragt [Le22]. Die Gespräche wurden als Tiefeninterviews persönlich mit Hilfe eines vorstrukturierten Fragebogens durchgeführt. Die Anzahl der zu diesem Zeitpunkt laktierenden Kühe unterlag einer großen Spannweite mit einem Minimum von 30 und einem Maximum von 230 Tieren. Die durchschnittliche Herdengröße lag insgesamt bei 116 Kühen, in den Betrieben in Rheinland-Pfalz lag sie bei 124 und in Luxemburg bei 85 Kühen.
Die eingesetzte Melktechnik auf den befragten Betrieben umfasste verschiedene Digitalisierungsstufen. Vier Betriebe melkten in einem Melkstand, in welchem keine oder fast keine Sensoren, wie z. B. nur ein Durchflusssensor, eingesetzt werden. Einen Melkstand, welcher auf drei teilnehmenden Betrieben genutzt wurde, ist mit mehreren Sensoren, wie beispielsweise zur Milchmengen-, Melkzeit, der Durchflussmengen- und der Leitfähigkeitsmessung ausgestattet. Mindestens einen und maximal vier Melkroboter hatten acht der teilnehmenden Betriebe im Einsatz. Mit Hilfe verschiedener weiterer Sensoren werden zudem Daten wie beispielsweise Temperatur, Inhaltsstoffe oder der Farbe gesammelt. Auf einem teilnehmenden Betrieb wurde mittels einer Rohrmelkanlage gemolken. Am häufigsten wurde insgesamt die Melktechnik der Marke GEA (https://www.gea.com) eingesetzt (31,25 %), gefolgt von den Marken DeLaval (https://www.delaval.com/), BouMatic (https://boumatic.com/), Miele und Fullwood (https://www.lemmer-fullwood.info). Die drei luxemburgischen Betriebe melken mit Melkrobotern und der Marke GEA (67 %) und DeLaval (33 %) (Abb. 3).

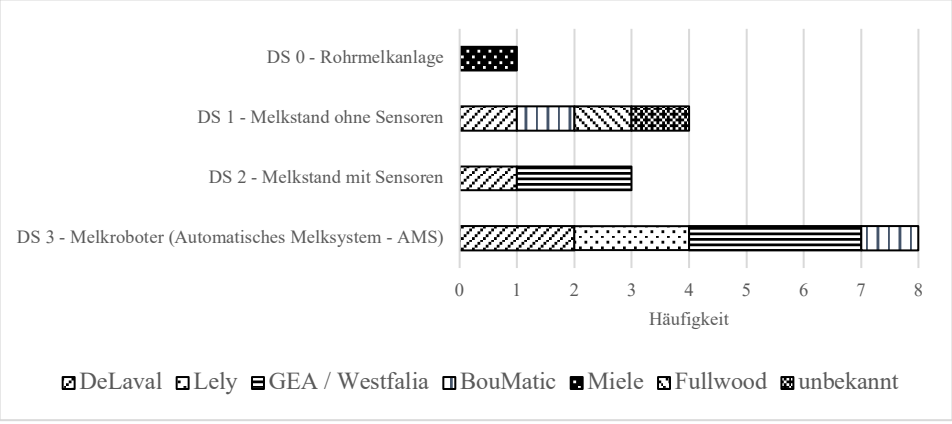


Abb. 3: Digitalisierungsstufe (DS) und eingesetzte Melktechnik bzw. -hersteller der teilnehmenden Betriebe [Le22].

Auch in den Bereichen der Fütterung, der Reproduktions- und Gesundheitsmanagement und des Datenmanagements wurde der Stand der Digitalisierung untersucht. Bei der

Fütterungstechnik verwendeten die meisten Betriebe die Total Mixed Ration (TMR). Einen selbstfahrenden Futtermischwagen verwendeten 20 % der Betriebe. Zur Unterstützung der Brunsterkennung, im Reproduktions- und Gesundheitsmanagement, nutzten zehn Betriebe Sensoren, die ein tierindividuelles Verhalten aufzeigen. Auf jedem dieser Betriebe wurde ein System zur Erkennung der Aktivität der Kuh eingesetzt. Zusätzlich nutzten sechs Betriebe Wiederkausensoren. Im Datenmanagement nutzen 93 % der Betriebe ein digitales Herdenmanagementsystem. Die Betriebe nutzten sowohl Kuhplaner als auch die Software der Melktechnikhersteller.
Zehn Befragte äußerten sich kritisch bezüglich der Schnittstellenproblematik. Die Insellösungen der einzelnen Systeme führten zu Mehrfacheingaben und einer komplizierteren Verwaltung der Daten. Sechs Befragte äußerten den Wunsch nach einem Herdenmanagementsystem, welches die Daten gut aufbereitet und interpretiert darstellt und zudem einfach zu bedienen ist. Ängste bezüglich der Datenhoheit, des Datenschutzes und der Datensicherheit wurden von fünf Befragten sowie Ängste bezüglich des Datenverlustes von drei Befragten geäußert. Drei Befragte verspürten durch die Digitalisierung eine höhere psychische Belastung, insbesondere hervorgerufen durch die notwendige permanente Erreichbarkeit für den Melkroboter. Des Weiteren wünschten sich drei Befragte eine bessere Internetverbindung und fallende Technikpreise. Gleichzeitig gab es Ängste bezüglich hoher Kosten durch Schäden und erschwerter Reparaturen, falls die Technik ausfallen sollte. Keine Ängste bezüglich der Digitalisierung äußerten drei Befragte. Weitere Probleme, die von einzelnen angesprochen wurden, waren die mangelnde Einarbeitung durch die Technikhersteller, die unklare gesellschaftliche Akzeptanz gegenüber weiterer Digitalisierungsschritte und der Wunsch nach mehr Einbeziehung der Praxis in die Entwicklung neuer Technologien.
Die Milchviehhaltung in Rheinland-Pfalz und in Luxemburg befindet sich in einem starken strukturellen Wandel. Mit den zunehmenden Herdengrößen geht die Automatisierung und Digitalisierung des Herdenmanagements fast zwingend einher. Diese qualitative Studie zeigt, dass die drei befragten luxemburgischen Betriebe auf einem gleichmäßig fortgeschrittenen Digitalisierungsstand sind. Die 12 rheinland-pfälzischen Betriebe befinden sich auf unterschiedlichen Digitalisierungsstufen in den einzelnen Bereichen der Milchviehhaltung. Insgesamt bietet die Digitalisierung viel Potenzial das Herdenmanagement zu unterstützen. Es besteht jedoch hoher Bedarf, die Befürchtungen und sehr konkreten Probleme der LandwirtInnen zu mindern, um den größtmöglichen Nutzen der Digitalisierung der Milchviehbetriebe zu erreichen.

3 Material und Methode

Zur Definition des Status-Quo der Digitalisierung in der Landwirtschaft wurden Befragungen in der Innen- und Außenwirtschaft durchgeführt. Als Konsequenz der durchgeführten Befragungen wurde der Aufbau eines Betriebsnetzwerkes als sinnvoll erachtet. Ein Aufruf nach Pilotbetrieben zur Teilnahme am Netzwerk wurde über die regionalen Fachjournale Rheinische Bauernzeitung (verbreitete Auflage 8.185 Exemplare [Me22]) und dem DLR-Wetterfax (etwa 700 Empfänger [No22]) verbreitet. Mit

Unterstützung des Bauern- und Winzerverbandes von Rheinland-Pfalz Süd e. V., ist inzwischen ein Netzwerk von etwa vierzig Pilotbetrieben mit unterschiedlichsten Produktionsrichtungen (die Pilotbetriebe decken alle gängigen Produktions-, Anbau- und Vermarktungsvarianten ab) und Digitalisierungsgraden (von Einsteiger bis Profi) sowie Ansprüchen an digitale Techniken entstanden.
Seit Beginn des Netzwerkes der Pilotbetriebe im Herbst 2020, ist die Intention, die Pilotbetriebe auf ihrem Stand der digitalen Technologien abzuholen und zu fördern. Für ein nachhaltiges Netzwerk bedarf es einer funktionierenden Kommunikation zwischen der Forschung, den Projektpartnern des Experimentierfeldes, den Behörden und dem Wissenstransfer zur Einbindung der digitalen Landwirtschaft in die Lehre (Abb. 1). In regelmäßigen Abständen von zwei Monaten, wird eine Online-Gesprächsrunde der Pilotbetriebe im Rahmen eines Themas zur Digitalisierung in der Landwirtschaft veranstaltet. Bei der Auswahl des Themas wird flexibel auf Themenvorschläge der Pilotbetriebe reagiert und vom Team des Forschungsprojektes Digitalisierung der TH Bingen entschieden. Die Gesprächsrunde wird im Online-Format mithilfe des Programms „MS-Teams“ (https://www.microsoft.com/de-de/microsoft-teams/log-in) durchgeführt. Die Teilnehmenden sind die Betriebsleiter der Pilotbetriebe, aber auch Studierende und weitere Interessierte werden eingeladen (über einen MS Teams-Link), an der Gesprächsrunde teilzunehmen. Die Gesprächsrunde beinhaltet Vorträge von jeweils 15 Minuten Dauer, mit anschließender Diskussionsrunde (jeweils 15 Minuten) und einer großen Diskussionsrunde von etwa 30 Minuten. Für eine Gesprächsrunde ist ein Zeithorizont bis zu zwei Stunden festgelegt. Im Anschluss der Veranstaltungen wird ein Protokoll erstellt und ein Artikel für die landwirtschaftlichen Zeitschriften, Rheinische Bauernzeitung und DWZ – Die Winzer Zeitung, verfasst. Abbildung 2 stellt den Prozessablauf des Informationsflusses einer Gesprächsrunde des Netzwerkes dar (Abb. 2).



Abb. 1: Aufbau eines nachhaltigen informellen Netzwerkes.

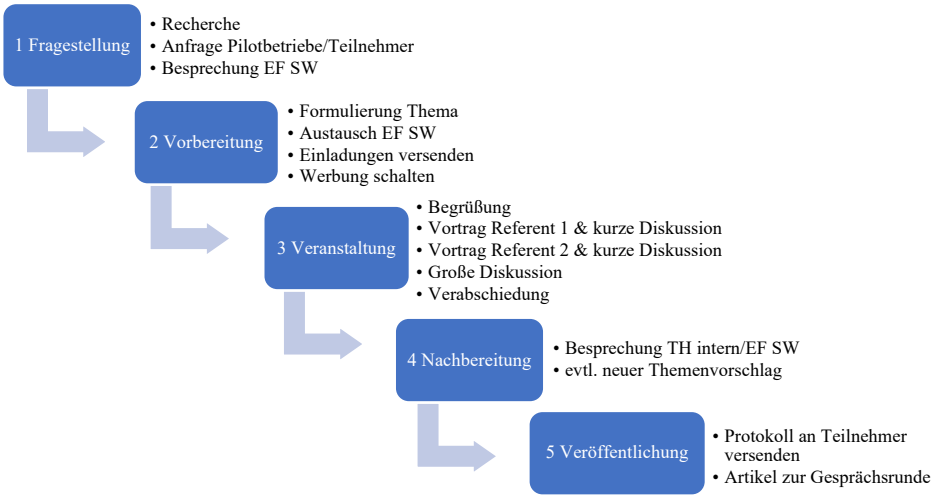


Abb. 2: Informationsfluss des Netzwerkes.

4 Ergebnisse

Nachfolgend werden die bisherigen Gesprächsrunden mit den Pilotbetrieben aufgelistet, welche von Februar 2021 bis März 2022 stattfanden und die daraus resultierenden Ergebnisse präsentiert. Der Teilnehmerkreis der Gesprächsrunden bestand aus den Pilotbetrieben, dem Team des Forschungsprojektes der TH Bingen und den eingeladenen Referenten. Außerdem wird die Wissenstransferstrategie zum Abbau der Hemmnisse gegenüber der Digitalisierung in der Landwirtschaft vorgestellt.

4.1 Erste Gesprächsrunde, Thema „Digitalisierung in der Landwirtschaft 1“

In den ersten zwei Gesprächsrunden der Pilotbetriebe wurden die DLG-Statements zur Digitalisierung in der Landwirtschaft [GUH18] zur Diskussion gestellt. Die DLG bezieht sich in den ersten Statements zur digitalen Landwirtschaft auf Themen, wie den Ausbau der Infrastruktur für die mobile Telekommunikation, die Datenhoheit des Landwirts und dem Ausbau von Datenschutz und Datensicherheit. Der Zustand der vorhandenen Infrastruktur (LTE, 5G) ist in weiten Teilen in Rheinland-Pfalz sowie in den Grenzregionen zu Luxemburg und Frankreich, ungenügend [Sp21]. Ein weiteres zentrales Statement der DLG ist die Datenhoheit des Landwirts. Nur der Landwirt kann den Zugriff auf seine erhobenen Betriebsdaten einem Dritten erlauben. Um sich nicht an einen Landtechnikhersteller binden zu müssen, wurde von den Pilotbetrieben das Interesse an herstellerunabhängigen Lösungsansätzen geäußert. Teilnehmende wiesen auf Beratungsbedarf zum Thema Digitalisierung in der Landwirtschaft hin, der von der Offizialberatung offensichtlich aktuell nicht gedeckt wird. Um die Betriebsdaten der Landwirtschaft schützen zu können, müssen Datenschutz und Datensicherheit ausgebaut werden. Grundsätzlich besteht keine Ablehnung, eine hohe Transparenz gegenüber Dritten, zu ermöglichen.

4.2 Zweite Gesprächsrunde, Thema „Digitalisierung in der Landwirtschaft 2“

Als Fortsetzung wurden in der zweiten Veranstaltung weitere Statements der DLG zur digitalen Landwirtschaft zur Diskussion gestellt [GUH18]. Das vierte Statement der DLG beinhaltet, dass die Landwirte aus den Geschäften mit ihren Daten Nutzen generieren sollten. Bisher generieren die Hersteller aus den Daten der Landwirte einen Nutzen und verwenden diese Daten zur Weiterentwicklung der Maschinen, wovon der Landwirt ebenfalls indirekt profitiert. Bei dem Kauf einer Maschine wird nicht danach gefragt, ob die Daten, welche in Zukunft mit der Maschine anfallen, vom Hersteller genutzt werden dürfen, es wird vorausgesetzt. Somit gibt es keine Entscheidungsmöglichkeit über Weitergabe der eigenen Daten, der Kauf ist die Einwilligung zur Datenweitergabe.
Das fünfte Statement der DLG fordert, dass mehr Transparenz bei der Big-Data-Analyse geschaffen werden sollte. Die Analyse der Daten des Landwirts könnte diesem bei der Entscheidungsunterstützung helfen (z. B. beim Maschinen- oder Betriebsmittelkauf).

In dem sechsten Statement der DLG wird der freie Zugang zu behördlichen Daten gefordert. Die Pilotbetriebe diskutierten über die GeoBox-Infrastruktur (http://geobox-infrastruktur.de/) von Rheinland-Pfalz [Ge21a]. Der GeoBox-Viewer (https://geobox-i.de/GBV-RLP/) ist Teil der Infrastruktur und bietet georeferenzierte Informationen wie Liegenschaftskarten, Luftbilder, topographische Karten und Verkehrswege zur Ansicht an [Ge21b]. Die Pilotbetriebe äußerten Bedenken darüber, dass diese landwirtschaftlichen Informationen für Jeden zugänglich sind. Ein Missverstehen dieser Informationen, von Personen ohne landwirtschaftliches Verständnis, könnten zu Komplikationen führen. Nach Meinung der Pilotbetriebe sollten die Daten des GeoBox-Viewers nur für Fachpersonal, über einen Login, einsehbar sein. Aufgrund des erhöhten Diskussionsbedarfs, stand das Thema der nächsten Gesprächsrunde schnell fest.

4.3 Dritte Gesprächsrunde, Thema „Vorstellung der GeoBox-Infrastruktur von Rheinland-Pfalz

Referenten vom Dienstleistungszentrum Ländlicher Raum (DLR) in Bad Kreuznach, Daniel Eberz-Eder und Lukas Wald, stellten die GeoBox-Infrastruktur von Rheinland-Pfalz vor. Die GeoBox ermöglicht den landwirtschaftlichen Betrieben den herstellerunabhängigen Austausch von notwendigen Daten mit beteiligten Akteuren. Somit bleibt die Datenhoheit bei den landwirtschaftlichen Betrieben. Das Ziel der resilienten GeoBox-Infrastruktur ist eine dezentrale Datenhaltung und regionale Vernetzung [Di22].
Die GeoBox-Infrastruktur enthält die digitalen Anwendungen des GeoBox-Viewers [Ge21b] und des GeoBox-Messengers [Ge21c]. Im GeoBox-Viewer werden georeferenzierte Informationen aber auch zeitkritische Daten, wie Wetterdaten, zur Verfügung gestellt. Ergänzend dazu, bietet die Plattform Daten zu Bodenarten, zur Erosionsgefährdung und der Ausweisung von spätfrostgefährdeten Flächen an [Di22].
Der GeoBox-Messenger ist eine digitale Plattform, welche als modernes Beratungs- und Kommunikationsmedium zur Vernetzung zwischen Beratern und Landwirten, aber auch als Kommunikation innerhalb eines landwirtschaftlichen Betriebes genutzt werden kann. Mit der Einrichtung von fachspezifischen Gruppen im GeoBox-Messenger, können kulturspezifische Fachinformationen schnell mit den Landwirten kommuniziert werden. Der GeoBox-Viewer stellt ausschließlich öffentlich zugängliche Daten zur Verfügung. Es wäre aber denkbar, dass bestimmte Bereiche im GeoBox-Viewer in Zukunft nur noch passwortgeschützt zur Verfügung stehen werden. Dennoch ist eine Arbeitserleichterung von Seiten der Pilotbetriebe spürbar, notwendige Daten (etwa für den Flächenantrag) wären schnell einsehbar und für verschiedene Fachbereiche (Bienen, Gemüsebau, Hausgarten, Landentwicklung, Obstbau, Pflanzenbau, Weinbau) verfügbar. Aufgrund der vielen Nachfragen, von Seiten der Pilotbetriebe, zum Thema Datenschutz in der Landwirtschaft, wurde das Thema der nächsten Gesprächsrunde bestimmt.

4.4 Vierte Gesprächsrunde, Thema „Datenschutz, Datensicherheit und Datenhoheit des Landwirts“

Die Digitalisierung in landwirtschaftlichen Betrieben führt dazu, dass vielfältige Daten aus unterschiedlichen Produktionsbereichen erfasst, gespeichert und verfügbar gemacht werden. Allerdings bringt die umfangreiche Datenerfassung auch Risiken hinsichtlich des Datenschutzes mit sich. Weiß ein landwirtschaftlicher Betrieb, wo sich seine betrieblichen Daten befinden? Welche Geschäftsdaten sind Betriebsgeheimnisse? Wer hat Zugriff auf Betriebsdaten und wem nutzen diese? Wie geht der Landwirt mit der Forderung nach mehr Transparenz um? Mit Unterstützung von zwei Juristen, Prof. Dr. Dieter Kugelmann, Datenschutzbeauftragter des Landes Rheinland-Pfalz und Prof. Klaus Gennen, LLR Rechtsanwälte und TH Köln, welche auf das Thema Datenschutz und Datensicherheit spezialisiert sind, konnten die Pilotbetriebe Fragen zu diesen Themen stellen. Von Seiten der Pilotbetriebe wurde angesprochen, dass Fehlinterpretationen von öffentlich verfügbaren Daten ein Problem sei, da hier Dritte falsche Schlüsse (ggf. auch öffentlich in der Presse) ziehen könnten. Die Juristen antworteten, dass Fehlinterpretation von Daten ein Kommunikationsproblem sei und nicht den Datenschutz betreffe, es jedoch einen Berichtigungsanspruch bei personenbezogenen Daten gäbe. Hier sollten ggf. äußerungsrechtliche bzw. presserechtliche Ansprüche durchgesetzt werden, je nach Art der Verunglimpfung. Es gilt: Je weniger Daten preisgegeben werden, desto weniger kommt es zu Fehldeutungen der Daten durch Dritte [Cl21]. Allerdings muss die Sensibilität gegenüber Daten und deren Schutzbedarf gesteigert werden. Hinsichtlich der Verwendung von öffentlich zugänglichen Daten im GeoBox-Viewer konnten die Referenten erklären, dass gesetzliche Grundlagen notwendig sind, um Daten in die Öffentlichkeit zu überführen. Ohne Rechtsgrundlage dürfen, von Seiten des Staates, keine Daten an die Öffentlichkeit gelangen. Die Pilotbetriebe stellten die Beziehung von Technik und Betriebsmitteln heraus, da es, aufgrund fehlender personenbezogener Daten, wenig Handlungsspielraum für den Landwirt gäbe. Bei der Nutzung von landwirtschaftlichen Maschinen zeichnen diese Daten auf (ähnlich wie beim PKW). Fahrgeschwindigkeit, Kraftstoffverbrauch, Drehzahlen und vielen weitere Daten werden erhoben und können in einer Fachwerkstatt ausgelesen werden, um z. B. Auskunft über das Fahrverhalten des Fahrers geben zu können. Diese Daten werden vom Fahrer erzeugt und können geladen und genutzt werden, sind aber juristisch gesehen keine personenbezogenen Daten und können somit an den Hersteller weitergegeben und verarbeitet werden. Die Juristen empfahlen Vertragsschlüsse mit Herstellern und/oder Lohnunternehmen, wodurch im Voraus geklärt wird, was nach Abschluss der Lohnarbeiten mit den Daten geschieht. Als problematisch stellt sich hierbei die asymmetrische Verhandlungsmacht heraus. Dies bedeutet, dass die Vertragspartner eine unterschiedlich gewichtete Bedeutung bei den Verhandlungen haben. Während der Gesprächsrunde wurde deutlich, dass die Landwirte, Winzer und Lohnunternehmer sich differenziert mit der Thematik Datenhoheit und Datensicherheit in der digitalen Landwirtschaft auseinandersetzen [WS21a].

4.5 Fünfte Gesprächsrunde, Thema „Digitalisierung auf Sonderkulturbetrieben“

Um den unterschiedlichen Stand der Digitalisierung auf Betrieben zu berücksichtigen, wurde eine Gesprächsrunde zu dem Thema „Digitalisierung in Sonderkulturbetrieben – Effizienzsteigerung und Arbeitszeiterfassung“ veranstaltet. Referent Lars Zimmermann, vom DLR Rheinpfalz (Forschungsanstalt für Weinbau, Phytomedizin, Oenologie und Gartenbau; https://www.dlr.rlp.de/DLR-RLP/Ueber-uns/DLRRheinpfalz), beschäftigt sich mit den digitalen Ansätzen, um die Arbeitseffizienz im rheinland-pfälzischen Obstbau zu steigern. Mit der Vorstellung von verschiedenen Obstbaukarteien, unteranderem auch Start-Up-Unternehmen konnte den Pilotbetrieben eine Übersicht über aktuelle verfügbare Programme zur Arbeitsunterstützung gezeigt werden. Der zweite Referent, Simon Orlob, teilte den Pilotbetrieben seine Erfahrungen als IT-Berater, zur Etablierung von digitalen Systemen anhand eines Arbeitszeiterfassungssystems (eigene Programmierung, selbst entwickeltes Programm) auf einem Erdbeeranbaubetrieb mit und berichtete wie die Digitalisierung eines landwirtschaftlichen Betriebes gut funktionieren kann. Mit der Vorstellung des Arbeitszeiterfassungssystems sprachen die Pilotbetriebe die Übertragbarkeit dieses Systems auf andere Betriebsformen an, da die Arbeitszeiterfassung ein wichtiges Thema auch für Weinbaubetriebe wäre. Bedenken äußerten die Pilotbetriebe bei den vielen Schnittstellen zu anderen Programmen, welche häufige Fehlerquellen darstellen können. Es wurde deutlich, dass der Stand der Digitalisierung auf Sonderkulturbetrieben noch nicht so weit vorangeschritten ist wie auf Markfruchtbetrieben oder viehhaltenden Betrieben [Ke21]. Eine der Ursachen ist, dass betriebsspezifische digitalen Anwendungen für Sonderkulturbetriebe nur sehr begrenzt auf dem Markt sind.

4.6 Sechste Gesprächsrunde, Thema „Wetterdaten und Digitalisierung in der Landwirtschaft“

Die Kenntnis der Wetterprognose oder der regelmäßig zu erwartenden Witterungsbedingungen sind in der Landwirtschaft von entscheidender Bedeutung, um optimale Erträge erzielen zu können. Mithilfe von Vorhersagen kann eine sinnvolle Planung von zum Beispiel Aussaat- und Ernteterminen, Pflanzenschutz-, Düngungs- und Bewässerungsmaßnahmen aufgestellt werden. Durch Wettervorhersagen können oft noch rechtzeitig Schutzmaßnahmen vor Frost, Hagel oder Sturm getroffen werden. Mit dem agrarmeteorologischen Messnetz von Rheinland-Pfalz ist eine standortgenaue Wettervorhersage möglich [Ag22] (https://www.dlr.rlp.de/Agrarmeteorologie). Durch die Erhebung sämtlicher Parameter (Windgeschwindigkeit in 2,5 m Höhe [m/sec], relative Luftfeuchte [%RF], Lufttemperatur ventiliert in 2 m Höhe [°C], Lufttemperatur ventiliert in 0,2 m Höhe [°C], Bodentemperatur in 0,05 m Tiefe [°C], Bodentemperatur in 0,2 m Tiefe [°C], Globalstrahlung [W/m²], Blattbenetzung [%], Niederschlag [mm]) [Ag22] erhalten die Praktiker zu der Wettervorhersage auch weitere ackerbauliche Entscheidungsunterstützungen. Durch die Veränderung des Klimas werden standortbezogene Wettervorhersagen für die zukünftige landwirtschaftliche Produktion immer wichtiger. Kleinräumiger Starkregen und trockenere Sommer werden zunehmen.

Dadurch wird die klimatische Wasserbilanz, also die Wasserversorgung an einem Standort, eine zentrale Fragestellung in der zukünftigen Landwirtschaft [St21a]. Mithilfe von Bodenfeuchtesensoren an verschiedenen Dauermessstellen in Rheinland-Pfalz, wird der neu modellierte Bodenfeuchtestatus von landwirtschaftlichen Flächen stündlich im GeoBox-Viewer veröffentlicht und dient damit als weitere Entscheidungsunterstützung [Ak22]. Die standortbezogene Interpretation der Wetterdaten bietet weitere Möglichkeiten der Entscheidungsunterstützung [Kl21].
Kurze Zeit nach dem Online-Workshop wurde durch die Flutkatastrophe in Teilen Deutschlands schmerzlich deutlich, dass bestehende Messtechniken und ein funktionierendes Messnetz deutlich ernster genommen werden müssen und genutzt werden sollten, um auf kurzfristig eintretende Naturkatastrophen frühestmöglich reagieren zu können. Die kontinuierliche Erfassung umfangreicher Wetterdaten ist weiterhin notwendig um langfristige Veränderungen richtig und standortbezogen zu verstehen und Anpassungsmöglichkeiten an den Klimawandel zu erproben [Kl21].

4.7 Siebte Gesprächsrunde, Thema „Low-Cost Apps für Marktfruchtbetriebe und viehhaltende Betriebe“

Verschiedene Pilotbetriebe machten die Angabe, dass sie Interesse an kostengünstigen und herstellerunabhängigen Lösungen zur Digitalisierung im Ackerbau und in der Viehhaltung hätten. Die verschiedensten digitalen Programme oder Apps sind heutzutage fester Bestandteil des Arbeitsalltages von Landwirtinnen und Landwirten. Aus dem Überangebot von digitalen Technologien ein gutes, praktisches und kostengünstiges Programm auszuwählen und zu testen, ist schwer und mühsam. Der Mitarbeiter, Christoph Weiß, vom Hofgut Neumühle in Münchweiler an der Alsenz (https://www.hofgut-neumuehle.de/), referierte über die Erstellung von Applikationskarten mithilfe der App „Solorrow“ (https://www.solorrow.com/). Die App ist eine kostengünstige Möglichkeit für einen landwirtschaftlichen Betrieb, eine Applikationskarte zu erstellen. Ohne in weitere Technik investieren zu müssen, kann die erstellte Applikationskarte manuell abgearbeitet werden. Die Betriebsleiterin eines Pilotbetriebes, Inse-Marie Stalter vom Wahlerhof Holstein GBR, stellte verschiedene digitale Anwendungen für ein Herdenmanagement vor. Den Kuhplaner „QMilch“ der QSX Datenschmiede GmbH ist eine Herdenmanagementsoftware zur Planung und Dokumentation eines Milchviehbetriebs (https://milch.qsxrind.com). Mit der App „Cowmanager“ wird jede Kuh im Betrieb rund um die Uhr überwacht (www.cowmanager.com). Die App „Netrind-Mobil“ (www.vit.de) umfasst Zuchtdaten, wie beispielsweise die Exterieureinstufungen, Abstammungen und Zuchtwerte. Das Programm der Vereinigten Informationssysteme Tierhaltung (vit) ist eine optimale Ergänzung zu QMilch. Die Schnittstellenproblematik der Einzelprogramme führt allerdings noch zu großer Unzufriedenheit in der Anwendung. Trotzdem bleibt die Arbeitserleichterung durch digitale Technologien im Vordergrund [WS21b].
Für die Nachverfolgung von Maschinen eines landwirtschaftlichen Betriebes stellte die Firma „exatrek“ (https://exatrek.de/) ihr Produkt durch Referent Dietrich Kortenbruck vor. Das Telemetriemodul von exatrek wird zurzeit von bestimmten Zielgruppen (wie

Großbetriebe und Lohnunternehmen) in Anspruch genommen. Einige Anwendungen haben Programme die teilweise auch auf PC oder Tablet dargestellt werden. Allerdings ist gerade bei Smartphone basierten Anwendungen, die Arbeit direkt an der Kuh, auf dem Acker oder an der Maschine ein großer Vorteil. Die präzise Dokumentation erhebt und analysiert Daten sehr genau. So werden Arbeitszeit, Spritverbrauch, aber auch Leerfahrten oder Bedienfehler transparent. Es ist z. B. auch ein CO_2-Ausstoß je Schlag ermittelbar. Ein wesentlicher Erfolgsfaktor der Anwendungen ist die Motivation der Nutzer. Dies gilt genauso für herstellerunabhängige, wie für herstellerbasierte Anwendungen, wobei eine weitergehende Betreuung des Herstellers nach der getätigten Investition ist weiterhin wichtig [WS21b].

4.8 Achte Gesprächsrunde, Thema „Low-Cost Apps für Sonderkulturbetriebe“

Nachdem die vorangegangene Veranstaltung sich mit kostengünstigen und herstellerunabhängigen Lösungen zur Digitalisierung im Ackerbau und in der Viehhaltung beschäftigt hatte, sollte die nächste Veranstaltung die Produktionsrichtungen Wein- und Obstbau behandeln. Anbieter verschiedener Apps und Praktiker aus den Bereichen des Wein- und Obstbaus stellten digitale Programme und Apps vor, sprachen über ihre praktischen Erfahrungen und standen zur Beantwortung von Fragen zur Verfügung. „Farmable“ (https://farmable.tech/de/) ist eine App für den Obstbau, welche Daten über den Weg erfasst, analoge Daten analysiert und neue Agrartechnologien (Wetterdaten, Sensoren, etc.) integriert. Daten und Analysen Dritter (Prognosen, Datenbanken, Spritzcomputer) werden mit einbezogen. Der Nutzen der App liegt darin, dass Erkenntnisse vorliegen und so Entscheidungen schnell getroffen werden können. „Vinumcloud“ (https://vinumcloud.de/) ist eine App für den Weinbau. Vinumcloud funktioniert über Tracking. Die Aktivitäten im Außenbetrieb werden in Echtzeit aufgezeichnet, Arbeitsfortschritte können verfolgt werden. Es sind historische Auswertungen möglich [WS22].
Zum Thema Schadensmanagement gibt es die Möglichkeit Schäden direkt vor Ort punktgenau per App zu melden. Die Maßnahmen zur Beseitigung der Schäden werden per App geplant und organisiert. Die Auftragsverwaltung geschieht über das Anlegen digitaler Aufträge und die anschließende Zuordnung zu einem bestimmten Mitarbeiter. Eine automatische Erkennung der Schläge lässt Aktivitäten „live“ mitverfolgen und so die Abarbeitung der Aufträge transparent werden. Die Teilnehmenden haben Fragen zur Flexibilität und Anpassungsfähigkeit der verschiedenen Programme und zur Beständigkeit von digitalen Nachweisen in Betriebskontrollen gestellt. Beide Vertreter der Programme gaben an, dass die Programme auf die Anforderungen der Kunden angepasst werden könnten. Eine Zeitersparnis durch Unterstützung in der Dokumentation konnten beide Programme nachweisen [WS22].

4.9 Neunte Gesprächsrunde, Thema „Digitale Bodenanalyse im Feld - der Weg in die Zukunft?“

Aufgrund der Bestimmungen und Anforderungen der aktuellen Düngeverordnung (DüV) wird das Wissen über den tatsächlichen und standortgenauen Bodenzustand für den Praktiker immer wichtiger, um die entsprechende Düngebedarfsermittlung auszuweisen [Ve17]. Nach bekundetem Interesse der Pilotbetriebe stellte Firma Stenon (https://stenon.io/) ihr Produkt FarmLab vor. Mit einem Sensormessstab („FarmLab“) soll eine digitale Bodenanalyse auf dem Feld, also in Echtzeit, vom Anwender durchgeführt werden. Nach Klärung der Fragen zum allgemeinen Gebrauch von FarmLab, wie zum Beispiel, in welcher Bodentiefe gemessen wird (16,5 cm Sensorkopf; Kalibration ist so vorgesehen, das Referenzwerte von 0 bis 30cm bereitgestellt werden), wie lange ein Messvorgang dauert (etwa dreißig Sekunden) und Kosten, stellten die Pilotbetriebe Nachfragen bezüglich der Datenhoheit. Die Ergebnisse der Bodenanalyse gehören dem Anwender, die Rohdaten (Spektraldaten), gehören Firma Stenon, welche primär zur Weiterentwicklung des Gerätes verwendet werden. Die Analysen von Firma Stenon sind nur online durchführbar. Neben Fragen zur naturwissenschaftlichen Methodik, wurden Fragen zum Datenschutz gestellt. Das Gerät zur Durchführung von digitalen Bodenanalysen steht nach den bisherigen Erkenntnissen in der Diskussion [Ol21] [St21b]. Die Messungen werden bisher nicht im Rahmen der Düngeverordnung anerkannt. Der Hersteller bemüht sich um eine Zulassung des Systems, durch die DLG, um die Methodik im Rahmen der DüV anwenden zu können. Eine Zulassung lag zur Zeit der Gesprächsrunde noch nicht vor. Nach dem Diskussionsverlauf zu urteilen, ist das Gerät noch weiter in der Entwicklung und größere, nachvollziehbare Praxistests unter kontrollierten Rahmenbedingungen und valide, öffentlich zugängliche Vergleichsuntersuchungen wären sicherlich wünschenswert. Es herrscht Weiterentwicklungsbedarf. Es gilt noch Potentiale der Digitalisierung zu erschließen.

4.10 Zehnte Gesprächsrunde, Thema „Arbeitsplatz Traktor – wie digital wird meine Bedienplattform?“

Der Arbeitsplatz Traktor steht vor dem Hintergrund der zunehmenden Digitalisierung im Wandel. Durch die Verwendung von Farmmanagementinformationssystemen (FMIS), wie zum Beispiel das John Deere Operation Center (https://operationscenter.deere.com/), das Fendt One (https://www.fendt.com/de/smart-farming/fendtone) oder das AFS Connect von Case IH (https://www.caseih.com/emea/de-de/produkte/afs%C2%AE-advanced-farming-systems/afs-connect), kommt die Entlastung des Fahrers zunehmend aus dem Büro des Betriebsleiters. Die auszuführenden Arbeiten werde vom Betriebsleiter mit einem FMIS im Büro geplant, der erstellte Auftrag an den Traktor gesendet, vom Fahrer durchgeführt und anschließend als „erledigt“ eingetragen. Weiterhin bleibt die größte Herausforderung die Vernetzung der Maschinen unterschiedlicher Hersteller untereinander [Wi22]. Zunächst wurden kritische Fragen bezüglich der Praktikabilität von autonom fahrenden Maschinen geäußert. Nach Aussage der Hersteller, sollen autonome Maschinen in Großbetrieben außerhalb von Europa zum Einsatz kommen. In Westeuropa

werden sich eher aufgerüstete Systeme durchsetzen [Mi22]. Der Arbeitsplatz eines Landwirts und damit auch die Bedienplattform hängt heute schon davon ab, ob der Betriebsleiter seinen Betrieb vom Büro führt, oder selbst die Arbeit auf dem Traktor erledigt [Ho22]. Die Arbeitszeit wird sich durch die ständige Überwachung durch autonome Maschinen nicht verringern, aber die Flexibilität der Gestaltung der Arbeit wird zunehmen [Pi13].

4.11 Wissenstransfer

Damit digitale Technologien in der landwirtschaftlichen Praxis zu ihrem vollständigen Nutzen gelangen, ist eine Steigerung der Fähigkeiten von Forschern, Anbietern von Software und Technologie und letztendlich von Landwirten und Beratern von hoher Bedeutung [IM20]. Um einen Transfer von Forschungsergebnissen in die landwirtschaftliche Praxis leisten zu können, müssen neue Wege geschaffen werden [JH19]. Mit Hilfe einer Wissenstransfer-Struktur können Erkenntnisse aus der Forschung in die Praxis vermittelt werden. Durch die Zusammenarbeit der Teams der Experimentierfelder sowie Beratung, Forschung und Praxis wurde zur Digitalisierung in der Landwirtschaft die Wissenstransferplattform FarmWissen (https://farmwissen.de/) erstellt. Wie die Ergebnisse der genannten Veranstaltungen deutlich machen, ist die Akzeptanz von digitalen Technologien von verschiedenen Hemmnissen abhängig. Dies spiegelt sich ebenfalls in der Literatur wider. Nach Schleicher und Gandorfer 2018 [Sc18], sind die größten Hemmnisse bei der Digitalisierung eine Inkompatibilität zwischen den Systemen, ein hoher Investitionsbedarf und Bedenken bei der Datenhoheit und -sicherheit. Diesen Hemmnissen möchte die FarmWissen-Strategie entgegenwirken. Durch eine unabhängige Vermittlung von fachspezifischem Wissen und die eindeutige Darstellung des Nutzens von digitalen Technologien und Anwendungen, wird eine individuelle einzelbetriebliche Förderung möglich. Anhand von Praxisbeispielen werden praktische Fragestellungen ausführlich beantwortet und der Nutzen von digitalen Technologien aufgezeigt. Die FarmWissen-Strategie beinhaltet unteranderem das FarmWiki (https://farmwiki.de/). Dort werden Glossareinträge von digitalen Begriffen in der Landwirtschaft zur Verfügung gestellt. Für komplexe Fragestellungen, unterstützt von Bild- und Videomaterial, können Tutorials aufgerufen werden. Die gewonnenen Forschungsergebnisse werden in der OpenDataFarm (https://odf.ef-sw.de/) visualisiert. Die OpenDataFarm dient der Verknüpfung von öffentlichen und betrieblichen Daten. Um den Wissenstransfer in der Praxis etablieren zu können, muss bereits in der Lehre begonnen werden. Mit Hilfe von frei zugänglichen Lernmanagementsystemen (LMS), wie zum Beispiel OpenOLAT (https://olat.vcrp.de/dmz/), wurden Online-Kurse mit Übungen konzipiert. Um Inhalte zur Digitalisierung in der Landwirtschaft im Unterricht und Vorlesungen einzusetzen.

5 Diskussion

Die Förderung der Einrichtung von Experimentierfeldern zur Digitalisierung in der Landwirtschaft des Bundesministeriums für Ernährung und Landwirtschaft hat sich als sinnvoll erwiesen. Erst mit der bundesweiten Auslegung des Förderprojektes von vierzehn Experimentierfeldern ist ein Austausch über Bundeslandgrenzen hinaus möglich. Im überregionalen Austausch entstehen neue Ansichtsweisen und Kontakte [Di21].
Die Akzeptanz von digitalen Technologien seitens landwirtschaftlicher Betriebsleiter korreliert positiv mit der Betriebsfläche sowie der gehaltenen Tierzahl in bzw. von Landwirtschaftsbetrieben [PT17]. Besonders in kleinstrukturierten Agrarlandschaften verläuft die Adoption von digitalen Technologien eher zögerlich [GG20]. Dennoch kann die Digitalisierung auch Chancen für kleine und mittelgroße Betriebe bedeuten [Sn17].
Die Ergebnisse der studentischen Arbeiten spiegeln sich in den Gesprächsrunden der Pilotbetriebe wieder. Die Pilotbetriebe geben positive Aspekte der Digitalisierung an, wie zum Beispiel die Arbeitserleichterung, Kosteneinsparung und verbesserte Umweltbedingungen an, allerdings scheinen die Herausforderungen der Digitalisierung, vor allem für Betriebe, welche in den Anfängen der Digitalisierung stehen, zu überwiegen. Mobilfunknetzabdeckung, Datenschutz und -sicherheit, Datenhoheit, Schnittstellenproblematik und die fehlende Betreuung des Herstellers sind wiederkehrende Themen in den Gesprächsrunden.
In der für Rheinland-Pfalz typischen Agrarstruktur spielt die Schweine- und Geflügelhaltung eine untergeordnete Rolle. Aufgrund dieser Besonderheit hat der Experimentierfeld Südwest diese Bereiche landwirtschaftlicher Produktion nicht bearbeitet. Dennoch wären Befragungen in diesen Produktionsbereichen interessant.

6 Schlussfolgerung

Das implementierte System der sich regelmäßig wiederholenden Gesprächsrunden eines Netzwerkes von Pilotbetrieben ist noch an einem sehr frühen Punkt der Erprobung. Zur Verstetigung und zur Eigendynamik sind weitere Optimierungen notwendig. Eine kontinuierliche Analyse des Systems wird auch weiterhin dazu führen, dass das beschriebene System sich zu einer nachhaltigen Plattform der Digitalisierung entwickelt. Durch die regelmäßigen Angebote, alle zwei Monate, zu den Gesprächsrunden der Pilotbetriebe wird im Verbundprojekt Experimentierfeld Südwest die Kommunikation zwischen den verschiedenen Akteuren gefördert. Anbieter von digitalen Technologien und Anwendungen stellen sich den Fragen, Wünschen und Erfahrungen der Pilotbetriebe und können Anregungen zur Verbesserung ihres Produktes aufnehmen. Die Pilotbetriebe profitieren von dem Erfahrungsaustausch mit den anderen Pilotbetrieben und bekommen die Chance Informationen über aktuelle Themen zu erfahren und offene Fragen direkt an den Anbieter zu richten. Die Offizialberatung steht im direkten Austausch mit der Praxis und kann das Angebot der Beratung auf die Bedürfnisse und Anregungen der Betriebe ausweiten.

7 Zusammenfassung und Ausblick

In einer kleinstrukturierten Agrarlandschaft wie in Rheinland-Pfalz ist der Stand der Digitalisierung in der Landwirtschaft auf einem geringen Niveau. Dennoch bietet die Digitalisierung Chancen auch für kleine bis mittelgroße Betriebe. Um die Digitalisierung in die Praxis zu bringen und Hemmnisse der Praktiker gegenüber der Digitalisierung abzubauen, wurde vom Verbundprojekt Experimentierfeld Südwest eine Wissenstransferstrategie entwickelt. Die Wissenstransfer-Plattform FarmWissen möchte den Praktikern den Zugang zur Digitalisierung erleichtern. Angeboten werden Praxisbeispiele zur Beantwortung einer praktischen Fragestellung, welche durch Unterstützung von Glossarbeiträgen und Tutorials detailliert erklärt werden. Um auf direkte Fragestellungen der Praktiker eingehen zu können, hat die TH Bingen ein Betriebsnetzwerk initiiert. Als Vorbereitung auf die Gründung des Netzwerkes wurden Befragungen zur Erfassung des aktuellen Standes der Digitalisierung in der Landwirtschaft durchgeführt. Mit dem Angebot von regelmäßigen Gesprächsrunden zu Themen rund um die Digitalisierung wird auf akute Fragestellungen seitens der Praktiker reagiert und ein einfacher Einstieg in die Digitalisierung aufgezeigt. Die Erkenntnisse der Gesprächsrunden werden in einschlägigen Medien publiziert und fließen mit in die Gestaltung der Vorlesungen.
Im weiteren Verlauf des Projektes werden Feed-Back-Studien durchgeführt, deren Ergebnisse zur Optimierung des Betriebsnetzwerkes genutzt werden. Für einen überregionalen und bundesweiten Informationsaustausch soll die Wissenstransfer-Plattform FarmWissen dienen. Die Herausforderung bleibt weiterhin, FarmWissen in der Praxis zu etablieren.

8 Danksagung

Die Förderung des Vorhabens „Verbundprojekt: Förderung des branchenübergreifenden und überbetrieblichen Datenmanagements zur Unterstützung landwirtschaftlicher Wertschöpfungssysteme (EF Südwest) - Teilprojekt 7“ erfolgt aus Mitteln des Bundesministeriums für Ernährung und Landwirtschaft (BMEL) aufgrund eines Beschlusses des deutschen Bundestages. Die Projektträgerschaft erfolgt über die Bundesanstalt für Landwirtschaft und Ernährung (BLE) im Rahmen der Förderung der Digitalisierung in der Landwirtschaft mit dem Förderkennzeichen 28DE111G18.
Ein Dank geht an die Projektleitung des Forschungsprojektes Digitalisierung der TH Bingen, Prof. Dr. Thomas Rademacher und an das Team des Forschungsprojektes: Prof. Dr. Clemens Wollny, Johannes Kunz, Kristin Leiendecker, Linda Visser, Jona Hinze und Hannah Spingat.

Literaturverzeichnis

[Ag22] Anonym: Agrarmeteorologie Rheinland-Pfalz – Wo? Dienstleistungszentrum Ländlicher Raum Rheinhessen-Nahe-Hunsrück (DLR RNH), Stand: 23.03.2022,

https://www.am.rlp.de/Internet/global/themen.nsf/b81d6f06b181d7e7c1256e920051ac19/79b5190741086c98c12574b100265d1d/$FILE/Stationen.pdf.

[Ak22] Anonym: Aktuelle Wetterdaten Rheinland-Pfalz. Dienstleistungszentrum Ländlicher Raum in Rheinland-Pfalz (DLR RLP), Stand: 13.03.2022, https://www.dlr.rlp.de/Agrarmeteorologie.

[Bü19] Büscher, W.: Digitalisierung des Stalles – aktueller Stand und Perspektiven. Züchtungskunde, 91 (1), 2019, S. 35-44.

[Cl21] Clasen, M.: Über den Wert von Daten in der Landwirtschaft - Meine Daten gehören mir! Aber was sind sie eigentlich wert? Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2021, S. 61.

[Di21] Anonym: Digitale Experimentierfelder – ein Beitrag zur Digitalisierung in der Landwirtschaft. Bundesministerium für Ernährung und Landwirtschaft, Dezember 2021, Stand 22.03.2022, https://www.bmel.de/DE/themen/digitalisierung/digitale-experimentierfelder.html.

[Di22] Anonym: Digitales Agrar Portal Rheinland-Pfalz (DAP). Dienstleistungszentrum Ländlicher Raum in Rheinland-Pfalz (DLR RLP), Stand: 13.03.2022, https://www.dap.rlp.de/Digitales-AgrarPortal/DAP.

[Eb21] Eberz-Eder, D.; Kraus, M.; Menne, J.; Wintz, L.: Experimentierfeld Südwest – Förderung des branchenübergreifenden und überbetrieblichen Datenmanagements zur Unterstützung landwirtschaftlicher Wertschöpfungssysteme. Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Referate der 41. GIL-Jahrestagung, Bonn, 2021, S. 367-372.

[GG20] Gabriel, A.; Gandorfer, M.: Landwirte-Befragung 2020 – Digitale Landwirtschaft Bayern, Bayrische Landesanstalt für Landwirtschaft (Lfl), 2020.

[Ge21a] Anonym: GeoBox-Infrastruktur - Projekt zur betrieblichen Datenhaltung und regionalen Vernetzung in der Landwirtschaft. Dienstleistungszentrum Ländlicher Raum in Rheinland-Pfalz (DLR RLP), Stand: 22.03.2022, http://geobox-infrastruktur.de/.

[Ge21b] Anonym: GeoBox-Viewer – Hilfeseite. Dienstleistungszentrum Ländlicher Raum in Rheinland-Pfalz (DLR RLP), Stand: 22.03.2022, https://geoservice.rlp.de/portal/apps/storymaps/stories/6d8b5a68b6494f0687cc45cb6ef1f85a.

[Ge21c] Anonym: GeoBox-Messenger, die kostenfreie Fachanwendung für die Landwirtschaft. Dienstleistungszentrum Ländlicher Raum in Rheinland-Pfalz (DLR RLP), Stand: 22.03.2022, https://www.geobox-messenger.rlp.de/.

[GUH18] Griepentrog, H. W.; Uppenkamp, N.; Hörner, R.: Digitale Landwirtschaft – Ein Positionspapier der DLG. DLG e. V.,2018

[Hö22] Höke, O.: Arbeitsplatz Traktor – Wie digital wird meine Bedienplattform? Referent von CLAAS, Vortrag Gesprächsrunde der Pilotbetriebe am 18.02.2022.

[IM20] Ingram, J.; Maye, D.: What are the implications of digitalisation for agricultural knowledge? Frontiers in Sustainable Food Systems, No. 4, p. 66, 202.

[JH19] Jánszky, B.; Herrmann, F.: Auf Augenhöhe - Wissenstransfer zwischen Forschung und Praxis der ökologischen und nachhaltigen Land- und Lebensmittelwirtschaft (Teilschlussbericht 1 des BÖLW). Bund Ökologische Lebensmittelwirtschaft e.V. (BÖLW), Berlin, 2019.

[Ke21] Keicher, R.; Rauber, B.; Schwarz, H.-P.; Brunner, D.: Bericht aus dem Experimentierfeld DIWAKOPTER – Digitalisierung im Weinbau und Ackerbau unter Nutzung von Multikoptern, vernetzter Sensorik und satellitengestützten Kommunikationskanälen. Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Referate der 41. GIL-Jahrestagung, Bonn, 2021, S. 385-390.

[Kl21] Klesius, P.; Trapp, M.; Wald, L.; Wintz, L.; Strerath, P.: Wie werden Wetterdaten erfasst und was bringt es der Landwirtschaft? Rheinische Bauernzeitung (RBZ), Nr. 42 / 23. Oktober 2021, S. 12-13.

[Ku21] Kunz, J.: Befragung von Pilotbetrieben im Rahmen des Experimentierfeldes Südwest über den aktuellen Stand der Digitalisierung in der Außenwirtschaft in Rheinland-Pfalz. Bachelorarbeit vorgelegt am 13.08.2021.

[Le22] Leiendecker, K.: Wie ist es um die Digitalisierung in der Milchviehhaltung bestellt? Rheinische Bauernzeitung (RBZ), Nr. 5 / 5. Februar 2022, S. 33-34.

[Me22] Anonym: Mediadaten 2022 – Basismedien für Rheinland-Pfalz und Saarland. Landvolk-Verlag GmbH.

[Mi22] Mithoff, T.: Arbeitsplatz Landmaschine – Vor dem Hintergrund zunehmender Digitalisierung, Referent von John Deere, Vortrag Gesprächsrunde der Pilotbetriebe am 18.02.2022.

[No22] Nöth, U.: Auskunft per Email von Ulrich Nöth vom Dienstleistungszentrum Ländlicher Raum Rheinhessen-Nahe-Hunsrück (DLR RNH) vom 23.03.2022.

[Ol21] Olfs, H. W.: Was kann der digitale Spaten? Top agrar 12/2021, S. 61-64.

[Pi13] Pierpaoli, E.; Carli, G.; Pignatti, E.; Canavari, M.: Drivers of Precision Agriculture Technologies Adoption: A Literature Review. Procedia Technology 8 (2013), pp. 61-69.

[PT17] Paustian, M.; Theuvsen, L.: Adoption of precision agriculture technologies by German crop farmers. Precision Agric (2017) 18, pp. 701-716

[Ri18] Anonym: Bundesministerium für Ernährung und Landwirtschaft - Richtlinie über die Förderung der Einrichtung von Experimentierfeldern zur Digitalisierung in der Landwirtschaft. Bundesministerium der Justiz und für Verbraucherschutz, Bekanntmachung, veröffentlicht am 21. September 2018.

[Sc18] Schleicher, S.; Gandorfer, M.; Ruckelshausen, A.; Meyer-Aurich, A.; Borchard, K.; Hofacker, C.; Loy, J.; Schwerdtfeger, R.; Sundermeier, H. F.; Theuvsen, B.: Digitalisierung in der Landwirtschaft - Eine Analyse der Akzeptanzhemmnisse. Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Referate der 38. GIL-Jahrestagung, Bonn, 2018, S. 203-206.

[Sn17] Schneider, W.: Digitalisieren oder weichen? – Wo bleibt die Landwirtschaft? Landinfo 3/2017, S. 8-12, 2017.

[Sp21] Spellerberg, A.: Technische Rahmenbedingungen - Breitbandversorgung in den Siedlungsräumen. In: Spellerberg, Annette (Ed.): Digitalisierung in ländlichen und verdichteten Räumen, 2021, Verlag der ARL - Akademie für Raumentwicklung in der Leibniz-Gemeinschaft, Hannover, S. 12-24.

[St18] Anonym: Statistisches Jahrbuch, Statistisches Landesamt Rheinland-Pfalz, Dezember 2018.

[St20] Sturm, B.; Nasirahmadi, A.; Müller, S.; Kulig, B.: Smart Livestock Farming – Eine Bestandsaufnahme. Züchtungskunde, 92 (6), 2020, S. 433-450.

[St21a] Anonym: Studie der Strategischen Behördenallianz – Anpassung an den Klimawandel. Deutscher Wetterdienst (DWD), 2021, Stand: 03.03.2022, https://www.dwd.de/DE/presse/pressemitteilungen/DE/2021/20210826_pm_behoerdenallianz.html.

[St21b] Anonym: Stellungnahme - Wir zweifeln den Test an. Firma Stenon, Top Agrar 12/2021, S. 65.

[Ve17] Anonym: Verordnung über die Anwendung von Düngemitteln, Bodenhilfsstoffen, Kultursubstraten und Pflanzenhilfsmitteln nach den Grundsätzen der guten fachlichen Praxis beim Düngen (Düngeverordnung - DüV). Bundesministerium für Ernährung und Landwirtschaft, Stand: 10.03.2022, https://www.gesetze-im-internet.de/d_v_2017/D%C3%BCV.pdf.

[Wi22] Wilmer, H.: Der Arbeitsplatz „Traktor“ vor dem Hintergrund zunehmender Digitalisierung. Referent von profi, Vortrag Gesprächsrunde der Pilotbetriebe am 18.02.2022.

[WS21a] Wintz, L.; Strerath, P.: Was muss der Landwirt wissen? Rheinische Bauernzeitung (RBZ), Nr. 34 / 28. August 2021, S. 17-19.

[WS21b] Wintz, L.; Strerath, P.: Digital sein ohne hohe Investitionskosten? Rheinische Bauernzeitung (RBZ), Nr. 44 / 8. November 2021, S. 16-17.

[WS22] Wintz, L.; Strerath, P.: Digitalisierung im Weinbau – ohne hohe Investitionskosten? Die Winzer-Zeitschrift, Februar 2022, S. 33.

Effizienz und Nachhaltigkeit durch Green-IT: ein systematischer Literaturüberblick im Kontext der Klimakrise

Laura Buhleier, Patrick Gantner, Tobias Frey, Michael Bors, Marc-André Kaufhold und Christian Reuter[1]

Abstract: Die Klimakrise gehört zu den aktuell größten Herausforderungen unserer Zeit. Informations- und Kommunikationstechnologien (IKT) können hierbei eine entscheidende Rolle spielen. Einerseits, da eine Effizienzsteigerung von oder durch IKT zu einer klimafreundlicheren Nutzung beitragen kann, und andererseits, da IKT zu einem erhöhten Ressourcenverbrauch führen kann. Um diese Thematik weiter zu untersuchen, wird in dieser Arbeit eine systematische Literaturrecherche durchgeführt, um Herausforderungen und Potenziale in der Adressierung der Klimakrise durch eine effiziente und nachhaltige Entwicklung des IKT-Sektors zu analysieren. Die dabei untersuchte Literatur beinhaltet Herausforderungen wie zum Beispiel den hohen Energie- und Materialverbrauch der IKT-Geräte und Datenzentren sowie die entstehenden Entsorgungskosten und das Konsumverhalten der Nutzer*innen. Deswegen sollten die Nutzer*innen mehr Informationen zur Wartung, zum Kauf gebrauchter Geräte, und zum Recyceln/Entsorgen erhalten. Die Analyse lieferte aber auch viele Potenziale. Durch IKT können Effizienzsteigerungen in den Bereich Industrie, Landwirtschaft, Verkehr und Transport erreicht und auch umweltschädliche Geräte substituiert werden. Durch diese Potenziale kann es aber, wie in anderen Bereichen, zu einem Rebound-Effekt kommen.

Keywords: Green-IT, Nachhaltigkeit, Effizienz, Rebound-Effekt, Klimakrise

1 Einleitung

Die Klimakrise ist eine der aktuell größten Herausforderungen. Der 6. Beurteilungsbericht des „Intergovernmental Panel on Climate Change“ (IPCC) aus dem Jahr 2021 zeigt erneut, dass der Mensch einen großen Einfluss auf den Klimawandel hat [MA21]. Informations- und Kommunikationstechnologie (IKT) stellt hier sowohl eine Möglichkeit als auch eine Herausforderung dar. So bietet IKT zum einen Möglichkeiten „die Effizienz verschiedener Systeme zu verbessern“ [ZTB09, S. 2], „die Notwendigkeit nach physischem Material in einer gegebenen Aktivität zu eliminieren oder zu reduzieren“ [S. 2] sowie „die Wahrnehmung zu erhöhen und menschliches Verhalten zu ändern“ [S. 2]. Demnach könnten „technologische Durchbrüche“ [KH20, S. 13] eine der letzten Möglichkeiten sein, der Klimakrise zu entkommen. Hier spielt auch die Effizienz eine bedeutende Rolle, denn sie „ist im Prinzip bestimmt von Technologie und Politik“ [HK20, S.13]. Neben diesen positiven Aspekten entstehend durch die Digitalisierung aber auch Herausforderungen bzgl. der Datenhoheit [Li21] oder technische Abhängigkeiten [Ku22].

Überdies hat auch die Produktion und Verwendung von IKT selbst Einfluss auf den Kli-

[1] Technische Universität Darmstadt, Wissenschaft und Technik für Frieden und Sicherheit (PEASEC), Pankratiusstr. 2, 64298 Darmstadt, <nachname>@peasec.tu-darmstadt.de, https://orcid.org/0000-0003-1920-038X

doi:10.18420/inf2022_83

mawandel, da sie 2.1-3.9% der globalen Emission ausmacht [FR21] und einen wachsenden ‚CO2-Fußabdruck' hat. „[Die] Analyse zeigt, dass nicht alle Zusagen im IKT-Sektor ehrgeizig genug sind, um die Klimaziele zu erreichen" [S. 1]. Deshalb sollte auch in IKT mehr Aufklärung zum Einfluss von IKT auf die Nachhaltigkeit durchgeführt werden. In dieser Arbeit wird daher die Dualität in der Verwendung von IKT unter Berücksichtigung der nachhaltigen Effizienz in der Klimakrise untersucht. Dabei liegt der Fokus auf einem Beitrag zu folgender Frage: **Welche Herausforderungen und Potenziale hat IKT in Bezug auf die nachhaltige Effizienz in der Klimakrise?**

2 Methodik

Für diese Arbeit wurde eine sequenzielle Literaturrecherche durchgeführt und um referenzierte Veröffentlichungen ergänzt. Als zugrundeliegende Struktur findet das 5-Phasenmodell von Brocke et al. [VO09] Anwendung. Dieses unterteilt die Recherche in: Definition des Umfangs der Arbeit, Konzeption des Themas, Literaturrecherche, Literaturanalyse und -synthese und zukünftige Forschungsfragen. Zu Beginn wurde durch die Lektüre verschiedener Forschungsarbeiten „ein Verständnis der Thematik entwickelt" [S. 219], um Suchparameter für die systematische Literaturrecherche festzulegen. Somit wurden in einem ersten Rechercheschritt zehn Arbeiten konsultiert. Anschließend wurden die Datenbanken Google Scholar, Web of Science (WoS), EBSCOhost und die digitale Bibliothek der Gesellschaft für Informatik (dl.gi.de) genutzt. Diese wurden nach den Themen Rebound-Effekt, Green ICT, Donut-Ökonomie, Effizienz und IKT und Nachhaltigkeit und IKT durchsucht, welche in der initialen Recherche als für die Zielsetzung der Arbeit relevant identifiziert wurden.

Die Recherchearbeit wurde unter den ersten vier Autor*innen aufgeteilt. Die Begriffe „Green ICT", „Efficiency", „Sustainability" und „Rebound Effect" wurden als Suchbegriffe gewählt. Zur Sicherstellung der Aktualität der Ergebnisse wurden nur Arbeiten, welche ab 2010 veröffentlicht worden sind, berücksichtigt. Ältere Veröffentlichungen wurden nur einbezogen, wenn durch besonders häufige Zitationen eine besondere Relevanz für das Forschungsfeld erkennbar war. Zur Erhöhung der Qualität der einzelnen Suchen wurden mehrere Datenbankabfragen mit jeweils unterschiedlichen Suchbegriffen auf unterschiedlichen Suchmaschinen durchgeführt. Insgesamt wurden vier Suchen zu den Schwerpunkten Nachhaltigkeit, Effizienz, Green IT und Rebound-Effekt durchgeführt. Das Ergebnis jeder Schwerpunktsuche enthält zehn Publikationen. Diese setzen sich aus ausgewählten Veröffentlichungen aus Google Scholar (3), Web of Science (3), EBSCOhost (3) und der digitalen Bibliothek der Gesellschaft für Informatik (1) zusammen, sodass die 30 relevantesten Ergebnisse der jeweiligen Suchanfrage berücksichtigt wurden.

Für den Bereich des Rebound-Effekts wurde die Suchbegriffe „"rebound effect" AND („ict" OR „it") AND „sustainability" AND „efficiency" verwendet (im Folgenden Rebound-Suche). Hier wurden 89 Treffer (67 nach Deduplikation) zur genaueren Untersuchung herangezogen. Durch ein anschließendes Titelscreening reduzierte sich die An-

zahl auf 42, und durch das Abstractscreening auf 34 Veröffentlichungen. Von diesen wurden 10 besonders geeignete Publikationen unter Berücksichtigung der beschlossenen Verteilung ausgewählt.

Für den Suchbereich ökologische Nachhaltigkeit und IKT wurde der Suchterm „"Sustainability" AND ("ICT" OR "information and communication technology" OR "Information Technology") AND "environmental"“ verwendet. Der Begriff „environmental“ diente zur Einschränkung auf ökologische Nachhaltigkeit, da es zuvor einige Ergebnisse gab, bei denen der Fokus auf anderen Formen von Nachhaltigkeit lag. Nach Anpassung wurde daher der Suchterm „"environmental sustainability" AND ("ICT" OR "information and communication technology" OR "Information Technology")” verwendet. Durch die Spezifikation wurden in den vier Datenbanken jeweils zwei Suchanfragen durchgeführt. Bei der Betrachtung der 240 Suchergebnisse (2 x 4 x 30) gab es viele Überschneidungen. Das Titelscreening reduzierte die Auswahl auf 47 Ergebnisse nach Deduplikation. Durch das Abstractscreening wurden 29 Ergebnisse ausgewählt und daraus wiederum zehn ausgewählte Veröffentlichungen genauer betrachtet.

Für den Schwerpunkt Effizienz von/in IKT wurde grundsätzlich der Suchterm „("Efficiency" AND ("IT" OR "ICT" OR "Ecological" OR "Environmental"))“ verwendet. Ausnahme war hier die Datenbank der digitalen Bibliothek der Gesellschaft für Informatik, welche keine Ergebnisse für die Suchabfrage lieferte, was zu einer Kürzung des Suchterms führte. Da sich nach Titel-, Keyword- und anschließendem Abstractscreening nicht genügend passende Arbeiten fanden, wurde der Suchterm weiter spezifiziert und die 10 passendsten Veröffentlichungen in das Endergebnis aufgenommen.

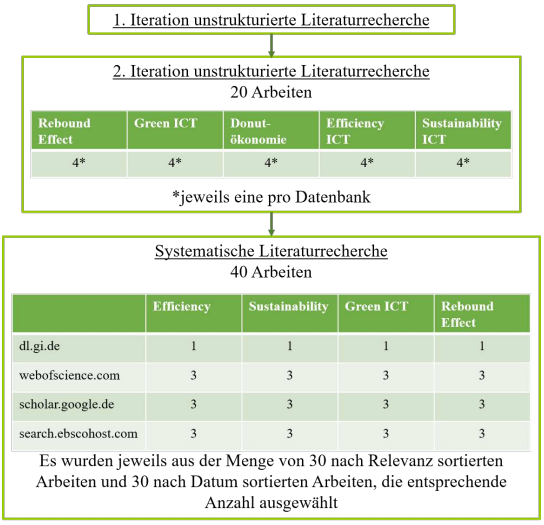


Abb. 1: Übersicht des Rechercheprozesses

Für den Suchbereich Green IKT wurde der Suchterm „("Green ICT" OR "Green Growth") & ("ICT" OR "IT" OR "Sustainability" OR "Efficiency")“ verwendet. Auch hier wurden Anpassungen vorgenommen, die schließlich zu 10 zentralen Arbeiten führten. Auffällig war, dass ein Großteil jener in der ersten Hälfte des Suchzeitraums 2010-2021 verfasst wurde. Eine Übersicht aller Suchanfragen lässt sich Abb. 1 entnehmen.

Insgesamt sind durch die durchgeführte Recherche 40 Werke in diese Arbeit eingegangen. Die Themengebiete umfassen Greening von (20) und Greening durch IKT (22), indirekte (15) und direkte Rebound-Effekte (15), Material- (3) und Energieeffizienz (11), CO_2-Einsparungen (9), Nachhaltigkeit (11) sowie die wirtschaftliche (10) und private Perspektive (4). Die Ergebnisse der Literaturrecherche werden nachfolgend beschrieben.

3 Ergebnisse

Frühere Arbeiten unterteilen die Auswirkungen von IKT in Primär-, Sekundär- und Tertiär-Effekte [VE17, HA14, CE14]. Primär-Effekte seien „Effekte, die durch die physische Existenz von IKT (Umweltauswirkungen durch Produktion, Nutzung, Recycling und Entsorgung von IKT Hardware) entstehen“ [HA14, S. 1]. Sekundär-Effekte seien „indirekte Auswirkungen auf die Umwelt von IKT aufgrund dessen Fähigkeit, Prozesse zu verändern (so wie Produktions- oder Transportprozesse), was in eine Modifikation (Abnahme oder Zunahme) ihrer Umweltbelastungen resultiert“ [HA14, S. 1]. Tertiär-Effekte wären dagegen „Umweltauswirkungen des Mediums- oder Langzeit-Adaption des Verhaltens (zum Beispiel Konsumverhalten) oder ökonomischer Strukturen durch die stabile Verfügbarkeit von IKT und der Services, die es bietet“ [HA14, S. 1]. Verdecchia et al. [VE17], Majeed [MA18], Aleksic [AL14] und Cecere et al. [CE14] verwenden äquivalente Definitionen, wobei letztere die genannten Unterteilungen teilweise auch als direkte, ermöglichende und systemische Effekte beschreiben. Mahdavi und Sojoodi [MS21] kontrastieren die Effekte zudem als positiv oder negativ.

	Positive Effekte	**Negative Effekte**
Primär- oder direkte Effekte	Die Nutzung von IKT in Umweltüberwachung und Kontrolle	Umweltauswirkungen der Herstellung und Nutzung von IKT-Geräten und -Ausrüstung
Sekundär- oder indirekte Effekte	Strukturelle Änderungen und Ressourcenschonung (eine Reduktion in der Menge der Ressourcen, die von der Wirtschaft verbraucht werden)	Unvollständige Substitution von Waren oder Services mit IKT
Tertiär-Effekte	Lifestyle Änderungen (z.B. Konsum auf einen grünen und umweltfreundlichen Konsum ausrichten)	Systemische oder Rebound-Effekte

Tab. 1: Darstellung der Primär-, Sekundär- und Tertiär-Effekte in Bezug auf deren positive und negative Auswirkungen, übernommen und übersetzt aus: Mahdavi und Sojoodi [MS21]

3.1 Greening von IKT

Nach Aleksic trug IKT in 2020 mehr als 4% zu den weltweiten CO2-Emissionen bei und „ungefähr 51% des IKT-bezogenen Energieverbrauchs [wurde] von der Telekommunikations-Infrastruktur und Datenzentren verbraucht“ [AL14, S. 1]. Nach Gorbenko et al. seien „die Informations- und Kommunikationstechnologien für ungefähr 2-2,5% der globalen CO2-Emissionen verantwortlich“ [GO18, S. 3] und „die gesamte Energieversorgung des gesamten Computing- und Kommunikationsequipment der Welt benötigt 160 GW pro Jahr, was etwa 8% der gesamten generierten Energie weltweit sind“ [S. 3]. Für eine Minderung der Umweltbelastung wäre also eine Steigerung der Effizienz in der Verwendung von IKT-Geräten notwendig.

Viele Autor*innen kritisieren an dieser Stelle die Ausrichtung der IKT-Forschung. So würden viele Forschungsarbeiten vor allem dem Energieverbrauch von Datenzentren und den während der Nutzung eines Gerätes entstehenden Umweltschäden Beachtung schenken und dabei die Auswirkungen von den Geräten der Endnutzer*innen und der Produktionsemissionen missachten [HA14B]. Es wird kritisiert, dass „die meisten der bisher absolvierten Studien sich ausschließlich auf die Energie konzentrieren, die während der Betriebs- (Nutzungs-) Phase verbraucht wurde“ [AL14, S. 4]. So könnten elektrische Autos zwar auf den ersten Blick für mehr Nachhaltigkeit sorgen, doch bei Einrechnung der Art der Stromgewinnung und den verwendeten Materialien würde ersichtlich, dass elektrische Autos nicht zwangsläufig für mehr Nachhaltigkeit sorgen [GO18]. Weiter wird kritisiert, dass in Studien zum Thema Green IKT die Rebound-Effekte wenig beachtet seien und der Fokus zu sehr darauf gelegt würde, energieeffizientere IT zu schaffen [HLW11]. Vergangene Entwicklungen würden jedoch zeigen, dass bessere Energieeffizienz zu einem insgesamt größeren Verbrauch durch mehr Nutzung führe.

Sollen kumulierte Folgen von IKT auf das Klima betrachtet werden, gibt es viele verschiedene Punkte zu beachten. So würde die Nachhaltigkeit von IKT-Geräten vor allem durch deren Lebensdauer bestimmt [AL14]. Aleksic [AL14] und Welfens und Lutz [WL12] schlagen daher „Life Cycle Assessments“ (LCA) vor, bei denen die umweltbelastenden Effekte eines Produktes von Herstellung bis Entsorgung betrachtet werden. Nach Gorbenko et al. stellt auch die mangelnde Leistungsfähigkeit von Software – etwa durch Alterungseffekte − ein Problem dar [GO18]. Sie warnen ebenfalls, dass sich unzuverlässige Netzwerke, zu welchen auch das Internet gehöre, durch die Verwendung mehrerer Systeme oder Prozesse zur Erhöhung der Fehlertoleranz negativ auf die Energienutzung auswirken würden. Außerdem erhöhe sich durch vermehrte Nutzung des Internets als auch durch den Gebrauch von Tablets, Smartphones und „Internet of Things (IoT)“-Geräten die Kommunikation über das Internet [AL14]. Das benötigte Wachstum an Bandbreite stelle ebenfalls höhere Anforderungen an die benötigten Netzwerkgeräte.

Zur Erreichung einer Effizienzsteigerung stellen Gorbenko et al. [GO18] vier Prinzipien der "Green Economics" vor: „Reduce“, wonach „Ressourcen-/Materialnutzung, Energieverbrauch und Abfall-/Schadstoffemission" [S. 2] verringert werden sollten, „Increase“, was bezeichnet, dass „die Anzahl der produzierten Produkte/Services, deren Langlebigkeit

und Leistungs-/Qualitäts-Indikatoren pro Einheit an verwendeten Ressourcen/Materialien, Energy und verursachtem Abfall/Schadstoffen“ [S. 2] erhöht werden sollten, „Reuse and Recycle“, wonach Produkte mit dem Gedanken der Wiederverwendbarkeit und Recyclebarkeit produziert werden sollten und schließlich „Recover and Renew“, was die Verwendung Erneuerbarer Energien und die Wiederverwendung von Teilmaterialien eines gebrauchten Produktes einschließt. Sie stellen für die Beurteilung von Prozessen und Produkten eine Metrik vor, anhand derer jeweils zwei Inputs und Outputs betrachtet werden sollten: Ressourcen und Energie, sowie Ergebnis und Abfallstoffe [GO18].

Stobbe et al. argumentieren, dass „eine Kombination aus spezifischer Hardware- und Softwareentwicklung […] der Schlüssel zur Verbesserung“ [ST16, S. 15] ist. Danach schlagen Drangmeister et al. vor, während der Entwicklung von Software eine Methode zum Messen des Energieverbrauchs und der Energieeffizienz zu verwenden, welche von Anfang bis zum Ende in den Entwicklungsprozess eingebunden werden sollte [DR13]. Weiterhin könnten „signifikante Einsparungen im Energieverbrauch durch Netzwerk-Infrastrukturen erreicht werden […], wenn sowohl das Netzwerkkonzept als auch die Architektur des Netzwerkelements in Beachtung des Energieverbrauchs optimiert werden“ [AL14, S. 3]. So kann eine Erhöhung der durchschnittlichen Auslastung die Anzahl der Server sowie den Energieverbrauch verringern [Go18, S. 4], denn eine „Verbesserung der Energieeffizienz von Computer Software“ [S. 5] gegenüber der Hardware sei zentral. Dafür seien besonders die Messung des Energieverbrauchs, sowie eine wohlüberlegte Energieerneuerung und intelligente Zeitplanung notwendig. Bieser und Hilty schließen, dass die Emissionen von Treibhausgasen (THG) im IKT-Sektor selbst sinken [BH18]. Es existieren also diverse Ansätze zur Effizienzsteigerung mit IKT. Jedoch merken Bol et al. an: „Zwar steigt die Energieeffizienz mit den Generationen der Kommunikationstechnologie, doch gleicht sie den gestiegenen Datenverkehrsaufwand nicht aus, so dass der globale Fußabdruck zugenommen hat“ [BPD21, S. 3].

3.2 Greening durch IKT

Weiterhin existiert die Möglichkeit, dass „die Anwendung von IKT in den verschiedenen Wirtschaftssektoren eine erhebliche Steigerung der Energie- und Ressourceneffizienz ermöglicht“ [SPL20, S. 12]. So hätten „durch IKT ermöglichte Lösungen das Potenzial […] die jährlichen Emissionen um 14,2% [bis 16,2%] der prognostizierten Gesamtmenge [von] 2020 zu reduzieren“ [AL14, S. 1]. Weitere Autor*innen nennen ähnliche potenziell positive Auswirkungen von IKT auf Nachhaltigkeit. Durch die Einführung von E-Mails und Lösungsfindungen durch automatisierte intelligente Verfahren habe IKT beispielsweise schon zu einer höheren Nachhaltigkeit beigetragen [MS21]. Zudem könne sich Green IKT durch die Innovativität und das schnelle Wachstum neuer Technologien positiv auswirken [WL12]. Weiterhin hätte IKT durch die Interkonnektivität einen positiven Einfluss auf Produktion und Handel. Insgesamt offenbaren sich vier Hauptpunkte für den Einfluss von IKT auf Nachhaltigkeit: „Softwareentwicklung“ und „Technologischer Fortschritt“, die zu einer erhöhten Effizienz und Nachhaltigkeit führten, „das Wachstum des Internets“,

welches zu niedrigeren „Kommunikationskosten“ führte und die generell weite Verbreitung des Internets, dessen Verwendung aber gleichwohl zu einer Verbreitung umwelt(un)freundlicherer Lebensstile führen könnte. Weiter führe IKT zu erhöhter Produktivität und Bildung und damit zu wirtschaftlichem Wachstum, das sowohl zu einem stärkeren Wunsch nach Umweltschutz als auch zu einem erhöhten Druck wirtschaftlich starker Regierungen auf Regierungen mit geringeren Umweltstandards führt.

Zur Ermöglichung dieses Greenings durch IKT benennen Gorbenko et al. folgende Ziele: „Die Entwicklung eines skalierbaren Modells für die Vergrünung der städtischen Umwelt“ [GO18, S. 6] , „die Entwicklung von ICT Services (...) für smarte städtische Umwelt“[S. 6], „die Entwicklung des [...] Smart-Grid-Konzepts“ [S. 6], „die Entwicklung von Modellen und Techniken, um die Sicherheit und Resilienz des Smart Grids abzuschätzen und sicherzustellen“ [S. 6] und die „Werbung für Green-Culture über Social Networks“ [S. 6]. Weiterhin entwickelten Bokolo und Mazlina ein Modell, um Unternehmensstrategien auf „Green Process“ [BM16] zu überprüfen, damit die Umweltverträglichkeit sichergestellt wird. Rivera et al. prüfen verschiedene sekundäre Effekte der IKT auf ihre Umweltverträglichkeit [RI13] aufbauend auf Yi und Thomas, welche Potenziale von IKT anhand von „E-Commerce“ konkretisieren: Virtualisierung von Produkten, Digitalisierung von Informationen, Dematerialisierung des Transports, Verkleinerung von Lager-/Büroflächen und Verkürzung der Lieferketten [YT07]. Weitere Anwendungsfelder von IKT zur effizienten Reduktion von CO_2-Emissionen seien die Industrie sowie das Bauwesen [NI20]. Nüssel und Hofmann et al. sehen solche Potenziale ebenso für die Landwirtschaft [NÜ18; HRA21] und den Verkehrssektor spricht der abnehmenden Pendelverkehr durch digitale Arbeit, wobei mögliche Rebound-Effekte hier noch untersucht werden müssen [HRA21; SBB20]. Auch im Frachttransport kann IKT einen Beitrag zur CO2-Reduktion leisten [CH21]. Obwohl es deutliche Einsparungen bei den Transportkosten durch IKT-Einkaufsysteme geben könne [DO00], entgegnen Sui und Rejeski, dass „Just-in-Time-Lieferungen dazu führen, dass Lastwagen halbleer fahren und der elektronische Handel schnellere Transportarten begünstigt, was den Kraftstoffverbrauch erhöht“ [SR02, S. 155-163].

Bei der Literaturrecherche wurden weitere Studien identifiziert, die das Potenzial von IKT zur allgemeinen Senkung der CO2 Emissionen untersuchen. So sorge IKT in Industrieländern für eine Senkung der CO2-Emissionen [BH18; KSA20]. Beim Ausbau von IKT-Infrastruktur in Entwicklungsländern steige das Emissionslevel hingegen, bis ein bestimmter Schwellwert erreicht wird. Erst bei einer Weiterentwicklung der IKT-Infrastruktur über diesen Schwellwert hinaus sänken die CO2-Emissionen [AGS17]. Dieses Ergebnis wird in weiteren Studien von Asongu et al. [ALB17], Majeed [MA18], Akande et al. [ACP19] und Khan et al. [KSA20] bestätigt.

3.3 Der Einfluss der Nutzer*innen

Verschiedene Autor*innen benennen zudem die Nutzer*innen von IKT-Geräten als einen großen Faktor für den CO_2-Fußabdruck [RI13; ST16; BA19; SPL20; CAS21]. Cordella et al. fassen zusammen, dass bis zu 49% von IKT-Produkten unbenutzt aufbewahrt werden,

nachdem sie ihr Lebensende erreicht haben. Nur 36 % würden in einen zweiten Lebenszyklus über- und 15% würden zum Recycling abgegeben. Diese Arbeit fokussiert das LCA von Smartphones, da der Beitrag jener bald größer als der von Desktops, Laptops und Bildschirmen ist [CAS21]. So wird geschätzt, dass nach aktuellem Stand der Technik (durch fortschrittlicherer Recyclingtechnologien könnte die CO_2-Einsparung ca. 30% höher sein) einzig anschließendes Recycling von Smartphones den CO_2-Fußabdruck um 5% senken könne. Beim Kauf von wiederaufbereiteten Geräten könne der Fußabdruck sich sogar um die Hälfte reduzieren. Erhöhen der Austausch von Batterie (1%) und Bildschirm (9%) in den ersten zwei Nutzungsjahren zunächst den CO_2-Fußabdruck, führt eine verlängerte Nutzungsdauer durch einen Bildschirmaustausch schließlich jedoch zu einer Einsparung von bis zu 44% bei 4 Jahren, die von wirtschaftlichen Einsparungen für Konsument*innen begleitet wird. Die zunehmende Bildschirm- und Speichergröße von Smartphones sowie „deren zunehmende Marktdurchdringung und kurze Austauschzyklen“ [CAS21, S. 14] führten derzeit noch zu hohen Emissionen. Zudem würde „die Analyse von Materialeffizienzaspekten durch das Vorhandensein möglicher Zielkonflikte erschwert“ [CAS21, S. 14] und eine wirksame Umsetzung von Materialeffizienzstrategien hinge von Verhaltensaspekten und umfassenden Informationen der Nutzer*innen ab, welche schließlich ihr Verhalten nachhaltiger gestalten müssten. Dies bedarf mehr Informationen für Verbaucher*innen, wie Geräte effizienter verwendet, instandgehalten und nach dem Lebenszyklus richtig entsorgt werden können. Eine Materialeffizienzstrategie ist also neben dem Verhalten von Nutzer*innen auch von der Informationsverbreitung abhängig.

Als mögliche Einflussfaktoren auf das Verhalten gelten wirtschaftliche Erwägungen, Werte, Einstellungen und Normen, sowie praktische Aspekte des täglichen Lebens [BE08]. In der Studie Berkers werden diese Punkte auf das Thema „Rolle der IKT für das Energieverhalten der Haushalte Energieeffizienz zur Eindämmung der globalen Erwärmung“ [BA19, S. 2] angewendet. Mit Hilfe von Technologien wie intelligenten Stromzählern zur Optimierung des Stromverbrauchs könnten so etwa der Energieverbrauch von Haushalten in den USA bis zu 22% und in der EU bis zu 20% gesenkt werden. Bezüglich einer absoluten Entkoppelung von Umweltleistungen vom wirtschaftlichen Wachstum durch die Digitalisierung argumentieren Bol et al., dass die Unmöglichkeit der wirtschaftlichen Entkoppelung von Energieverbrauch und Treibhausgasen ein Grund der Zunahme der CO_2-Emissionen sei. So sei die Herausforderung, dass in unserem nicht-idealen Markt die „low-carbon“ Energie, welche Datenzentren verwenden, an anderen Orten fehle. Ihrer Meinung nach „bleibt die Entkopplung lokal, und auf globaler Ebene wird nur eine begrenzte Verringerung der Treibhausgasemissionen erreicht, da der Energie-Fußabdruck nicht kleiner wird“ [BPD21, S. 4]. Hinzu kommen die bereits diskutierten Anforderungen durch Nutzer*innen: Die hohe Austauschrate von Verbraucherendgeräten und die Schaffung künstlicher Bedürfnisse, wie 4K-Videostreaming auf Smartphones oder IoT-Anwendungen, welche auch als „Internet-of-Shit“ bezeichnet werden [BPD21]. Damit eine absolute Entkoppelung irgendwann erreicht werden könnte, müssten die Ressourceneffizienz erhöht und gleichzeitig die Rebound-Effekte limitiert werden [SPL20]. Durch die Komplexität der Diskussion ist aber nicht sicher festzustellen, ob die Digitalisierung die Entkopplung mildert oder verstärkt. Jedoch sollten Politiker*innen nicht darauf hoffen, dass die Digitalisierung von alleine zu einer nachhaltigen absoluten Entkopplung führe

[SPL20]. „Die Wahl und das Verhalten des Nutzers [bleiben also] ein Schlüsselelement im Hinblick auf das verbleibende Energiesparpotenzial“ [ST16, S. 6].

3.4 Rebound-Effekt

Neben den direkten Effekten von IKT durch Material- und Energiebedarf [SC13] entstehen auch indirekte Effekte, die die Anwendung von IKT und die daraus resultierenden Verhaltens- und Struktureffekte umfassen [RO12]. Eine einheitliche und vollständige Zuordnung dieser Effekte in primäre, sekundäre und tertiäre oder Effekte höherer oder niedrigerer Ordnungen ist in der Literatur jedoch nicht zu finden [PHF19]. Dies gilt besonders für Effekte wie den Rebound-Effekt, der sich aus den Wechselwirkungen zwischen technischen Systemen und Nutzerverhalten ergibt [HSA16] und Situationen bezeichnet, in denen berechnete Einsparpotenziale durch Effizienzsteigerungen schließlich nicht oder nur teilweise realisiert werden können, weil die Reduktion an der einen zu einem höheren Verbrauch an anderer Stelle führt. Galvin argumentiert, dass diese klassische Einteilung in direkt und indirekt in der Anwendung von IKT deshalb unpassend sei, da Verbraucherbedürfnisse hier nicht befriedigt, sondern stetig neu geschaffen würden [GA15]. Andere teilen speziell bei IKT-Anwendungen auftretende Rebound-Effekte in direkte und indirekte wirtschaftliche sowie zeitliche und räumliche Rebound-Effekte ein [RI13]. Weiterhin wurden bei der Analyse mehrerer Studien zu den Effekten von IKT-Anwendungen direkte Rebound-Effekte im Zusammenhang mit einer verstärkten Nutzung aufgrund von Kosten- und Zeiteinsparungen sowie durch unerwartetes Nutzungsverhalten festgestellt [PHF19]. Es wird außerdem auf die weniger erforschten Arten in Zusammenhang mit Zeit, Platz und Motivation hingewiesen [PHF19].

Diese verschiedenen Einordnungen und damit verbundenen abweichenden Definitionen gelten als Ursache für die stark voneinander abweichenden Studienergebnisse zur Stärke des Rebound-Effekts [GGD00]. Die Studien zum Rebound-Effekt fokussieren zumeist auf den Energieverbrauch und stehen nicht in direktem Zusammenhang mit IKT-Anwendungen [GGD00; SO07]. In der zuvor beschriebenen Literaturrecherche wurden jedoch auch Studien identifiziert, die sich speziell auf den Rebound-Effekt von IKT-Anwendungen fokussieren. Takahashi et al. berechneten den Rebound-Effekt von IKT-Diensten in einer Fallstudie über Videokonferenzen, in der der Rebound-Effekt die erreichten Kohlenstoffeinsparungen um bis zu 20% verringern kann [TA04]. Joyce et al. fanden Rebound-Effekte im Zusammenhang mit der IKT-Nutzung in Schweden von über 100% [JO19]. Ray Galvin kommt in seiner Studie zu Rebound-Effekten des IKT- und Elektronik-Sektors ebenfalls zu Rebound-Effekten von über 100% [GA15]. Eine Effizienzsteigerung führe demnach zu einer Erhöhung des Energieverbrauchs. Ein weiterer Effekt, der auch als Miniaturisierungsparadoxon bekannt ist und sich den direkten Rebound-Effekten zuordnen lässt, wurde von Hilty untersucht. Hier wurde festgestellt, dass sich die durchschnittliche physische Masse eines Mobiltelefons zwischen 1990 und 2005 zwar um den Faktor 4,4 verringerte, die Gesamtmasse aller Telefone in der Schweiz jedoch um den Faktor 8 zunahm. Dies ist darauf zurückzuführen, dass „Rechenleistung schneller billiger wird, als

dass sie kleiner wird“ [HI08]. Da es sich bei IKT häufig um Allzwecktechnologien handelt, kommen Einsparungen einem höheren Einkommen gleich, das in anderen Sektoren ausgegeben werden kann [MA11]. Dieser indirekte Rebound-Effekt wurde in einer Vielzahl von Studien festgestellt.

Im Bereich des E-Learnings haben Herring und Roy festgestellt, dass bei verschiedenen Systemen zur Bereitstellung von Bildungsmaterialien eine elektronische Bereitstellung im Vergleich zum druckbasierten Fernunterricht nicht zu einer Verringerung des Energieverbrauchs oder der CO2-Emissionen führt. Grund dafür sind Rebound-Effekte beispielsweise durch die Nutzung von Computern und der Heizung zu Hause [HR02]. Online Unterricht führe jedoch zu einer Dematerialisierung [CLS13]. Auch im Bereich der Telearbeit kommt es zu erheblichen Rebound-Effekten. So kann Telearbeit zu längeren Pendelwegen führen, wenn die physische Anwesenheit im Büro erforderlich ist, da die Arbeitnehmer sich möglicherweise entscheiden, weiter von ihrem Arbeitsplatz entfernt zu wohnen, wenn sie wissen, dass sie die Möglichkeit zur Telearbeit haben [GO15]. Diese indirekten Rebound-Effekte übersteigen im Zusammenhang mit der gestiegenen Kaufkraft und den erhöhten Reisezeiten die direkten Effizienzgewinne bei weitem [HR02]. Weitere Einsparungen im Transportbereich, beispielsweise durch Teleshopping, werden durch die Zunahme des Verkehrs für andere Zwecke wie weitere Einkäufe oder durch andere Familienmitglieder zu großen Teilen zunichte gemacht [FA12]. Auch die Verwendung immer effizienter werdender Navigationsgeräte kann zu Einsparungen auf einzelnen Strecken führen, führt jedoch auch dazu, dass insgesamt mehr Fahrten getätigt werden [GO15].

Nach Hakansson und Finnveden [2015] können bei Effizienzsteigerungen in der Produktion von IKT-Geräten starke Rebound-Effekte auftreten. Bei Effizienzsteigerungen in der Verwendung von IKT-Geräten sind auftretende Rebound-Effekte jedoch wesentlich geringer. Hier kann es zu tatsächlichen Einsparungen im Ausstoß von CO2-Emissionen kommen. Des Weiteren stellten Hakansson und Finnveden [2015] fest, dass erhöhte Investitionen in IKT-Produkte zwar einen Anstieg der Emissionen im IKT-Bereich zur Folge haben, diese aber einen „umgekehrten“ Rebound-Effekt aufweisen, also der Konsum in anderen Bereichen zurückgeht (Substitution). Außerdem fanden Hakansson und Finnveden [HF15] im Widerspruch zu Chitnis et al. [CH14] heraus, dass wohlhabendere Verbraucher stärkere Rebound-Effekte verursachen. Aus wirtschaftlicher Sicht ist ein hoher Rebound-Effekt jedoch wünschenswert, da Gewinne am höchsten sind, wenn der Rebound am höchsten ist [HI12]. Deshalb müssen nach Hilty zur Vermeidung eines Rebound-Effekts regulierende Maßnahmen getroffen werden, die einen Wachstum verhindern [HI12]. Er schlägt außerdem entgegen wirtschaftlicher Logik vor, nur Prozesse, deren begrenzender Faktor ein anderer als der Energieverbrauch ist, zu optimieren, da hier die Rebound-Effekte geringer sind. Ein solcher könnte zum Beispiel die benötigte Zeit sein. Andere betonen, dass technologischer Fortschritt wichtig sei, um Rebound-Effekte zu vermeiden und schlagen neben Energiepreismechanismen eine Umstrukturierung der Energiequellen vor [LW19]. So könne erneuerbare Energie dazu führen, den Rebound-Effekt abzuschwächen. Aus der neueren Literatur geht hervor, dass verschiedene wirtschaftliche Instrumente wie Umweltsteuern, die Bepreisung von Ressourcen oder die Fest-

legung von Grenzen für die Ressourcennutzung für die Vermeidung von Rebound-Effekten wirksam sein könnten [FV20].

4 Diskussion

Das IKT-Paradoxon beschreibt den Widerspruch, dass IKT einerseits negativ zur Klimakrise im Sinne von Ressourcennutzung, CO_2-Ausstoß und Energieverbrauch beitrage, andererseits jedoch durch intelligente Lösungen, eine Vereinfachung oder Erübrigung von Transportwegen und Energieeinsparungen hilfreich für die Bekämpfung der Klimakrise sein könne [QU19]. Die Betrachtung dieser Dualität ist deshalb relevant, weil „das IKT Paradoxon einige nützliche Einsichten in die Bewältigung von Umweltshocks des Klimawandels" [QU19, S. 626] bietet. Daher werden im Folgenden die Herausforderungen und Potenziale von IKT in Bezug auf die nachhaltige Effizienz in der Klimakrise diskutiert.

4.1 Herausforderungen und Potenziale von IKT in der Klimakrise

Vor allem unter dem Begriff Greening von IKT wurde festgehalten, dass IKT sich *negativ* auf das Klima auswirken kann. Demnach können sowohl Produktion, Wartung, Betrieb als auch Entsorgung und Recycling die Umweltbilanz von IKT negativ beeinflussen. Insgesamt sei IKT für ca. 2,1-3,9% der globalen Emissionen verantwortlich [FR21]. Ein Problem im Umgang mit aktuellen Zahlen zu den Umweltauswirkungen von IKT stellt dabei häufig die Fokussierung in Forschung und Datensammlung dar [HA14B, AL14, GO18, HA15]. Dabei würde sich etwa ausschließlich an den Betriebskosten orientiert und verwendete Materialien, Produktions- und Entsorgungskosten, sowie Rebound Effekte ignoriert [HA14B, AL14, GO18, HA15]. Auch die Effizienz von Geräten und Netzwerken hätten noch Verbesserungspotenziale [GO18, AL14]. Nach Bol et al. kann daher „die Verbesserung der Effizienz […] nicht das einzige ökologische Ziel sein, da sie eher zur Erzielung wirtschaftlicher Gewinne als zur Verringerung des globalen Fußabdrucks eingesetzt wird" [BPD21, S. 24]. Zusätzlich argumentieren sie, dass der zunehmende IKT-Wohlstand und das daraus folgende Wirtschaftswachstum nicht durch Innovation vorangetrieben werden sollten, weil dies oft nicht mit den Grenzen unseres Planeten vereinbar sei [BPD21]. Ebenso sei zu bedenken, dass die Entkoppelung durch die Digitalisierung, zumindest in den nächsten ein bis zwei Jahrzenten, noch verstärkt würde [SPL20]. Weiterhin stellt sich die Frage, ob die Verbraucher*innen ihr Verhalten zu mehr Nachhaltigkeit anpassen und dadurch auch mehr für nachhaltige Produkte und Services zahlen würden [ebd.]. Aktuell ginge die Tendenz eher zu Geräten, welche einen höheren CO_2-Fußabdruck haben [BPD21; CAS21]. Da ein hoher Rebound-Effekt aus wirtschaftlicher Sicht wünschenswert sei [HI12], bestünde eine besondere Herausforderung darin, diesen zu vermeiden. Derzeit kämen Einsparungen durch Effizienzsteigerungen dem Verbraucher als zusätzliches Einkommen zugute, das wiederum in anderen Bereichen ausgegeben werden kann, sodass Einsparungen in dem einen Bereich zu einem Konsumanstieg in einem anderen führen.

Weiter konnte festgestellt werden, dass sich IKT auch durchaus *positiv* auf die Klimakrise auswirken kann. Eine Lösung für die einseitige Kostenbetrachtung in Forschung und Datenerhebung könnte das „Life Cycle Assessment" darstellen, wonach die ganze Lebenszeit eines IKT-Produktes für die Evaluation beachtet wird. Ein Ansatz zur Effizienzsteigerung sind vor allem die von Gorbenko et al. eingeführten „Green Economics" sowie deren Herausforderungen Reduzierung, Erhöhung, Reuse und Recycle, sowie Recover und Renew [GO18]. Auch eine effiziente Zusammenarbeit von Software und Hardware sowie genaue Messungen der Umweltkosten könnten hier Abhilfe schaffen [ST16, AL14, GO18]. Ebenfalls sollten Modelle und Konzepte zur Effizienzsteigerung entwickelt und umgesetzt werden [GO18]. Weitere Potenziale von IKT liegen in ihrer Ausweitung. So könne IKT in anderen Disziplinen wie der Industrie, Landwirtschaft und Verkehr eine Effizienzsteigerung bewirken, die sich positiv auf Emissionen, Energie und Ressourcennutzung auswirkt [SPL20, AL14]. Auch die technischen Innovationen, die IKT mit sich bringt, könnten zu einer erhöhten Interkonnektivität und geringeren Transfer- und Kommunikationskosten führen [MS21, WL12]. Der Umwelteinfluss sollte ebenso eine Rolle beim Erstellen eines Produktes sein wie die funktionalen Leistungsmerkmale und das ästhetische Design [ST16]. Hierbei sollte den Nutzer*innen mehr Informationen geliefert werden, wie sie ihre Geräte effizienter nutzen können, um den Lebenszyklus eines Gerätes zu verlängern und dadurch den CO_2-Fußabdruck zu verringern. Genauso wird die Forschung vorangetrieben, neue, nachhaltigere, und besser recyclebare Materialen zu finden, um die Materialeffizienz jedes Geräts zu steigern. Rebound-Effekte bieten auch Potenziale zur Verbesserung der Umweltbelastung. So beschreiben Hilty et al. einen umgekehrten Rebound-Effekt, bei dem umweltfreundliche IKT-Produkte umweltschädliche Produkte aus anderen Bereichen substituieren [HLW11].

4.2 Zusammenwirkungen von Herausforderungen und Potenzialen

Einerseits kann eine höhere Effizienz in Industrie, Landwirtschaft Verkehr und Transport dazu beitragen, weniger Ressourcen zu verbrauchen und somit ökologisch nachhaltiger zu werden. Andererseits können die Folgen durch Rebound-Effekte und dem damit einhergehendem Wirtschaftswachstum zu einer noch stärkeren ökologischen Belastung führen. Negative Rebound-Effekte durch Substituierung von umweltschädlichen Produkten durch IKT-Produkte können dagegen zur ökologischen Nachhaltigkeit beitragen. Der Einfluss von IKT auf das Konsumverhalten der Nutzer*innen könnte hierbei eine Rolle spielen. Vor allem die Materialeffizienz ist hier in Verbindung zur Suffizienz und Konsistenz zu betrachten [HLW11]. Zum einen ist es wichtig, neue, nachhaltigere, besser recyclebare Materialen zur Herstellung der Geräte zu finden, aber zum anderen ebenso den Nutzer*innen den gesamten CO_2-Fußabdruck ihrer Geräte näher zu bringen.

Herausforderung	Potenzial
• Fokus von Forschung und Datenerhebungen • Produktions-, Wartungs-, Nutzungs- und Entsorgungskosten (Effizienz von Geräten) • Vermeidung von Rebound-Effekten • Konsumverhalten der Nutzer*innen	• Life Cycle Assessment • Green Economics, Zusammenarbeit von Software und Hardware, genaue Messungen, Modelle und Konzepte zur Effizienzsteigerung • Effizienzerhöhung in Industrie, Landwirtschaft, Verkehr und Transport • Interkonnektivität und geringere Transfer-Kommunikationskosten • Positive Auswirkung auf Produktivität, Bildung und Wirtschaftswachstum • Substituierung umweltschädlicher Produkte durch umweltfreundliche IKT-Produkte

Tab. 2: Herausforderungen und Potenziale von IKT

5 Fazit

In dieser Arbeit wurden die Herausforderungen und Potenziale von IKT in Bezug auf die nachhaltige Effizienz in der Klimakrise untersucht und dargestellt. Dabei wurden vor allem Green IKT und der Rebound-Effekt mit Blick darauf betrachtet, wie sich diese auf die Nachhaltigkeit und Effizienz auswirken. Die in der Literaturrecherche identifizierten Überthemen Greening durch IKT, Greening von IKT, Einfluss der Nutzer*innen und Rebound Effekt strukturierten folglich die Rechercheergebnisse, um die Herausforderungen und Potenziale von IKT und deren Wechselwirkungen darzustellen. Limitiert ist diese Recherche dabei durch die begrenzte Anzahl der untersuchten Artikel und die eingeschränkten Suchbegriffe. Insgesamt zeigt sich, dass es bislang an Studien und Betrachtungen von IKT in Bezug auf das Zusammenspiel der drei Nachhaltigkeitsstrategien Effizienz, Suffizienz und Konsistenz in der Klimakrise mangelt. Diese Arbeit stellt jedoch eine Vielzahl an Möglichkeiten der Effizienzsteigerungen von und durch IKT dar, die das enorme Potenzial von IKT zum Umweltschutz verdeutlichen.

Die großen Einflüsse von IKT und die komplexen Verstrickungen über alle drei Säulen der Nachhaltigkeit hinweg sind in ihrer Gesamtheit schwer zu erfassen. So führt technologischer Fortschritt als Ursache für Effizienzsteigerungen im IKT-Bereich in erster Linie zu einem Wirtschaftswachstum, dessen Nachhaltigkeit nicht zwangsläufig gegeben ist. Um diese zu gewährleisten, müssten Inputfaktoren begrenzt werden. Potenzielles Wachstum müsse also aufgrund von Effizienzsteigerungen durch Konzepte der nachhaltigen Suffizienz oder Konsistenz reguliert werden. Dieses Zusammenspiel der drei Nachhaltigkeitsstrategien sollte in zukünftigen Forschungsarbeiten genauer untersucht werden. Um die Chancen und Potenziale von IKT zu nutzen und die Herausforderungen zu bewältigen ist

also eine interdisziplinäre Zusammenarbeit notwendig. Somit bleibt es zukünftigen Forschungsarbeiten offen, aktuelle Erkenntnisse zu den dargestellten Themen einzuholen. Dabei kann diese Arbeit als Orientierung für zu beobachtende Herausforderungen und Potenziale und deren Wechselwirkungen dienen. Auch Weiterentwicklungen in den genannten Bereichen können auf diese Arbeit aufbauen.

Danksagung

Diese Arbeit wurde aus Mitteln des Zweckvermögens des Bundes bei der Landwirtschaftlichen Rentenbank im Rahmen des Projekts GeoBox II sowie AgriRegio gefördert.

Literaturverzeichnis

[ACP19] Akande, A., Cabral, P., und Casteleyn, S.: Assessing the Gap between Technology and the Environmental Sustainability of European Cities. Inf Syst Front 21/ 3, pp. 581–604, 2019

[AL14] Aleksic, S.: Green ICT for sustainability: A holistic approach. In 37th International Convention, pp. 426–431, 2014.

[AGS17] Añón Higón, D., Gholami, R., und Shirazi, F.: ICT and environmental sustainability: A global perspective. Telematics and Informatics 34/4, pp. 85–95, 2017.

[ALB17] Asongu, S., Le Roux, S., und Biekpe, N.: Enhancing ICT for Environmental Sustainability in Sub-Saharan Africa. SSRN Journal, 2017.

[BA19] Bastida, L., Cohen, J.J., Kollmann, A., Moya, A., und Reichl, J.: Exploring the role of ICT on household behavioural energy efficiency to mitigate global warming. Renewable and Sustainable Energy Reviews, 103, pp. 455–462, 2019.

[BE08] Berker, T.: Energienutzung im Heim als soziotechnische Praxis. In (Fischer, C., ed): Strom sparen im Haushalt. Oekom, München, pp. 175-192, 2008.

[BH18] Bieser, J., und Hilty, L.: Assessing Indirect Environmental Effects of Information and Communication Technology (ICT): A Systematic Literature Review. Sustainability 10/8, 2662, 2018.

[BM16] Bokolo, A.J., und Mazlina, A.M.: Development of a Green ICT Model for Sustainable Enterprise Strategy. JSCDSS, 3/3, pp. 1-12, 2016.

[BPD21] Bol, D., Pirson, T., und Dekimpe, R.: Moore's Law and ICT Innovation in the Anthropocene, pp. 19–24, 2021.

[RI13] Rivera, M., Håkansson, C., Svenfelt, Å., und Finnveden, G.: Including second order effects in environmental assessments of ICT. Environmental Modelling & Software 56, pp. 105–115, 2014.

[CLS13] Caird, S., Lane, A., und Swithenby, E.: ICTs and the Design of Sustainable Higher Education Teaching Models: An Environmental Assessment of UK Courses. In (CAEIRO,

S., FILHO,W.L., JABBOUR, C. UND AZEITEIRO, U.M: ED.), Sustainability Assessment Tools in Higher Education Institutions. Springer International Publishing, Cham, pp. 375–385, 2013.

[CE14] Cecere, G., Corrocher, N., Gossart, C., und Ozman, M.: Technological pervasiveness and variety of innovators in Green ICT: A patent-based analysis. Research Policy 43/10, pp. 1827–1839, 2014.

[CH21] Chatti, W.: Moving towards environmental sustainability: information and communication technology (ICT), freight transport, and CO2 emissions. Heliyon 7/10, e08190, 2021.

[CH14] Chitnis, M., Sorrell, S., Druckman, A., Firth, S.K., und Jackson, T.: Who rebounds most? Estimating direct and indirect rebound effects for different UK socioeconomic groups. Ecological Economics, 106, pp. 12–32, 2014.

[CAS21] Cordella, M., Alfieri, F., und Sanfelix, J.: Reducing the carbon footprint of ICT products through material efficiency strategies: A life cycle analysis of smartphones. Journal of Industrial Ecology 25/2, pp. 448–464, 2021.

[DO00] Downey, M.: Workshop on implications of the new digital economy on transportation: developing research and data needs, 2000.

[DR13] Drangmeister, J., Kern, E., Naumann Stefan, Sparmann, G., und Guldner, A.: Greening Software with Continuous Energy Efficiency Measurement, 2013.

[FA12] Falch, M.: Environmental Impact of ICT on the Transport Sector. In (HADJIANTONIS, A.M. AND STILLER, B. ed.) Telecommunication Economics, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 126–137, 2012.

[FV20] Freire-Gonzalez, J., und Vivanco, D.F.: Pandemics and the Environmental Rebound Effect: Reflections from COVID-19. Environ Resource Econ, 2020.

[FR21] Freitag, C., Berners-Lee, M., Widdicks, K., Knowles, B., Blair, G.S., und Friday, A.: The real climate and transformative impact of ICT: A critique of estimates, trends, and regulations. Patterns (New York, N.Y.) 2/9, pp. 0–18, 2021.

[GA15] Galvin, R.: The ICT/electronics question: Structural change and the rebound effect. Ecological Economics 120, pp. 23–31, 2015.

[GO18] Gorbenko, A., Tarasyuk, O., Kor, A.-L., und Kharchenko, V.: Green economics: A roadmap to sustainable ICT development. In (IEEE ed.) 9th International Conference on Dependable Systems, Services and Technologies (DESSERT), pp. 561–567, 2018.

[GO15] Gossart, C.: Rebound Effects and ICT: A Review of the Literature. In (HILTY, L.M. ed.) ICT Innovations for Sustainability, Springer International Publishing AG, Cham, pp. 435–448, 2015.

[GGD00] Greening, L.A., Greene, D.L., und Difiglio, C.: Energy efficiency and consumption — the rebound effect — a survey. Energy Policy 28/6-7, pp. 389–401, 2000.

[HF15] Hakansson, C. und Finnweden, G.: Indirect rebound and reverse rebound effects in the ICT-sector and emissions of CO2. Proceedings of EnviroInfo and ICT for Sustainability 2015. Advances in Computer Science Research, 66-73, 2015.

[HA14] Hankel, A.: Understanding Higher Order Impacts of Green ICT, 2014.

[HA14B] Hankel, A., Oud, L., Saan, M., und Lago, P.: A Maturity Model for Green ICT: The case of the SURF Green ICT Maturity Model, 2014.

[HR02] Herring, H., und Roy, R.: Sustainable services, electronic education and the rebound effect. Environmental Impact Assessment Review 22/5, pp. 525–542, 2002.

[HK20] Hickel, J., und Kallis, G.: Is Green Growth Possible? New Political Economy 25/4, pp. 469–486, 2020

[HI12] Hilty, L.: Why energy efficiency is not sufficient–some remarks on "Green by IT". In (Arndt, H.-K., Knetsch, G. und Pillmann, W., eds.): EnviroInfo Dessau 2012, Part 1: Core Application Areas. Shaker Verlag, Aachen, 2012.

[HLW11] Hilty, L., Lohmann, W., und Huang, E.M.: Sustainability and ICT - An overview of the field. Notizie di Politeia 27/104, pp. 13-28, 2011. https://www.zora.uzh.ch/id/eprint/55640/.

[HI08] Hilty, L.M.: Information technology and sustainability. Essays on the relationship between ICT and sustainable development. Books on Demand, Norderstedt, 2008.

[HA15] Hilty, L.M., und Aebischer, B.: ICT for Sustainability: An Emerging Research Field. In (HILTY, L.M. ed.) ICT Innovations for Sustainability, Springer International Publishing AG, Cham, pp. 3–36, 2015.

[HRA21] Hofmann, J., Ricci, C., und Ansu-Holz, D.: IT und Nachhaltigkeit – eine Einführung. HMD 58/1, pp. 6–23, 2021.

[HR02] Hoogeveen, M.J., und Reijnders, L.: E-commerce, paper and energy use: a case study concerning a Dutch electronic computer retailer. IJEP 18/6, 541, 2002.

[HSA16] Horner, N.C., Shehabi, A., und Azevedo, I.L.: Known unknowns: indirect energy effects of information and communication technology. Environ. Res. Lett. 11/10, 103001, 2016.

[JO19] Joyce, P.J., Finnveden, G., Håkansson, C., und Wood, R.: A multi-impact analysis of changing ICT consumption patterns for Sweden and the EU: Indirect rebound effects and evidence of decoupling. Journal of Cleaner Production 211, pp. 1154–1161, 2019.

[KSA20] Khan, F.N., Sana, A., und Arif, U.: Information and communication technology (ICT) and environmental sustainability: a panel data analysis. Environ Sci Pollut Res Int 27/29, pp. 36718–36731, 2020.

[Ku2] Kuntke, F., Linsner, S., Steinbrink, E., Franken, J., Reuter, C.:Resilience in Agriculture: Communication and Energy Infrastructure Dependencies of German Farmers. International Journal of Disaster Risk Science (IJDRS), 13, pp. 214–229, 2022.

[LW19] Liao, M., und Wang, L.: China's Energy Consumption Rebound Effect Analysis Based on the Perspective of Technological Progress. Sustainability 11/5, p. 1461, 2019.

[Li21] Linsner, S., Kuntke, F., Steinbrink, E., Franken, J., Reuter, C.: The Role of Privacy in Digitalization – Analysing the German Farmers' Perspective Proceedings on Privacy Enhancing Technologies. Proceedings on Privacy Enhancing Technologies (PoPETs), 3, pp. 334–350, 2021.

[MS21] Mahdavi, S., und Sojoodi, S.: Impact of ICT on Environment, 2021. PREPRINT (Version 1) available at Research Square [https://doi.org/10.21203/rs.3.rs-1020622/v1].

[MA18] Majeed, M.T.: Information and Communication Technology (ICT) and Environmental Sustainability in Developed and Developing Countries. Pakistan Journal of Commerce and Social Sciences 12/3, pp. 758–783, 2018.

[MA21] Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S.L., Péan, C., Berger, S., Caud, N., Chen, N. Goldfarb, L., Gomis, M.I., Huang, M:, Leitzell, K., Lonnoy, E., Matthews, J.B.R., Maycock, T.K., Waterfield, T., Yelekçi,O., Yu, R., und Zhou, B.: IPCC, 2021: Summary for Policymakers. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. In Climate Change 2021: The Physical Science Basis, IPCC, Cambridge University Press, 2021.

[MA11] Maxwell, D., McAndrew, L., Owen, P., Muehmel, K., und Neubauer, A.: Addressing the Rebound Effect. A report for the European Commission DG Environment, 2011.

[NI20] Nizam, H.A., Zaman, K., Khan, K.B., Batool, R., Khurshid, M.A., Shoukry, A.M., Sharkawy, M.A., Aldeek, F., Khader, J., und Gani, S.: Achieving environmental sustainability through information technology: "Digital Pakistan" initiative for green development. Environ Sci Pollut Res Int 27,/9, pp. 10011–10026, 2020.

[NÜ18] Nüssel, M.: Landwirtschaft 4.0 – die Waffe gegen Hunger und Umweltzerstörung? In (C. BÄR, T. GRÄDLER UND R. MAYR ED.) Digitalisierung im Spannungsfeld von Politik, Wirtschaft, Wissenschaft und Recht, Springer Gabler, Berlin, pp. 343–363, 2018.

[PHF19] Pohl, J., Hilty, L.M., und Finkbeiner, M.: How LCA contributes to the environmental assessment of higher order effects of ICT application: A review of different approaches. Journal of Cleaner Production 219, pp. 698–712, 2019.

[QU19] Qureshi, S.: Climate change adaptation for sustainable development: the information and communication technology (ICT) paradox. Information Technology for Development 25, 4, pp. 625–629, 2019.

[RO12] Røpke, I.: The unsustainable directionality of innovation – The example of the broadband transition. Research Policy 41, 9, pp. 1631–1642, 2012.

[SPL20] Santarius, T., Pohl, J., und Lange, S.: Digitalization and the Decoupling Debate: Can ICT Help to Reduce Environmental Impacts While the Economy Keeps Growing? Sustainability 12, 18, 7496, 2020.

[Sc13] Schien, D., Shabajee, P., Yearworth, M., und Preist, C.: Modeling and Assessing Variability in Energy Consumption During the Use Stage of Online Multimedia Services. Journal of Industrial Ecology 17, 6, pp. 800–813, 2013.

[So07] Sorrell, S.: The Rebound Effect: An Assessment of the Evidence for Economy-Wide Energy Savings from Improved Energy Efficiency. UK Energy Research Centre, 2007. https://ukerc.ac.uk/publications/the-rebound-effect-an-assessment-of-the-evidence-for-economy-wide-energy-savings-from-improved-energy-efficiency/

[SBB20] Stephan, B., Büttner, L., und Breitkreuz, A.: Arbeiten nach Corona. Warum Homeoffice gut fürs Klima ist. Werkstattbericht / IZT, Institut für Zukunftsstudien und Technologiebewertung 119. Greenpeace e.V, Hamburg, 2020.

[St16] Stobbe, L., Proske, M., Beucker, S., Hintemann, R., und Lang, K.-D.: Energy efficiency of ICT: Further improvement through customized products. In (IEEE ed.) 2016 Electronics Goes Green 2016+ (EGG), pp. 1–7, 2016.

[SR02] Sui, D.Z., und Rejeski, D.W.: Environmental impacts of the emerging digital economy: the e-for-environment e-commerce? Environmental management 29, 2, pp. 155–163, 2002.

[Ta04] Takahashi, K.I., Tatemichi, H., Tanaka, T., Nishi, S., und Kunioka, T.: Environmental impact of information and communication technologies including rebound effects. In (IEEE International Symposium on Electronics and the Environment ed.), Conference Record, pp. 13–16, 2004.

[Ve17] Verdecchia, R., Ricchiuti, F., Hankel, A., Lago, P., und Procaccianti, G.: Green ICT Research and Challenges. In (V. Wohlgemuth, F. Fuchs-Kittowski und J. Wittmann, Eds.) Advances and New Trends in Environmental Informatics, Springer International Publishing, Cham, pp. 37–48, 2017.

[Vo09] vom Brocke, J., Simons, A., Niehaves, B.B., Riemer, K., und Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process, 2009.

[WL12] Welfens, P.J.J., und Lutz, C.: Green ICT dynamics: key issues and findings for Germany. Miner Econ 24, 2-3, pp. 155–163, 2012.

[YT07] Yi, L., und Thomas, H.R.: A review of research on the environmental impact of e-business and ICT. Environment international 33, 6, pp. 841–849, 2007.

[ZTB09] Zapico, J.L., Turpeinen, M., und Brandt, N.: Proceedings of the 4th International Conference on Persuasive Technology. ACM Other conferences. ACM, New York, NY, 2009.

Datenqualität und Qualitätsmetriken in der Datenwirtschaft (DQ)

Application-specific quality metrics for the assessment of data for deep learning from large datasets

Gesa Marie Götte [1], Bonito Steffen Thielert [2] and Andreas Herzog [3]

Abstract: Application-specific quality metrics support getting suitable data from large databases to pre-train deep neural networks or getting good statistical measures. Especially when using high-dimensional or multimodal sensor data from industrial processes the small amount of training examples from each device or plant must be supplemented by additional data. We present a system for the definition of application-specific metrics in a model composed of statistical functions and neural networks. Further, we introduce a business model for using this system for the interaction of data providers with their customers. In order to obtain suitable data, the user sends his request to the data provider in the form of a quality metric model and gets back the best fitted data. Our system helps the user to define the model through examples and by setting the model parameters through genetic algorithms.

Keywords: application-specific quality metrics, transfer learning, provider customer relationship.

1 Introduction

Transfer learning is a suitable way to overcome the problems associated with small amounts of training examples, especially for high-dimensional or multimodal data [TSK18, MBS18]. This is used in the field of image or sound recognition, where deep learning networks, pre-trained with millions of images, can be easily transferred to your own application.

The first processing layers of deep learning models often extract general features from the data, while the later layers represent the dependencies between these features and task-specific-target values. The non-task-specific features from the first layers can be reused by integrating the already learned feature extraction in the learning process for the new application.

This principle will be transferred to industrial process data. There is often not enough training data available for individual applications. When machines are new or the data

[1] Fraunhofer IFF, Logistics and Factory Systems, Sandtorstrasse 22, 39106 Magdeburg Germany, gesa.goette@iff.fraunhofer.de
[2] Fraunhofer IFF, Logistics and Factory Systems, Sandtorstrasse 22, 39106 Magdeburg Germany, bonito.thielert@iff.fraunhofer.de
[3] Fraunhofer IFF, Logistics and Factory Systems, Sandtorstrasse 22, 39106 Magdeburg Germany, andreas.herzog@iff.fraunhofer.de, https://orcid.org/0000-0003-4084-6003

doi:10.18420/inf2022_84

collection has just started, the amount of data is rather small. This is especially challenging when the target attribute is highly unbalanced as for rare events. Additionally, the industrial process data are often very high-dimensional. Therefore, it can be necessary to use additional training data from other sources and related scenarios to enable a successful learning process with suitable outcomes. Selecting appropriate training data to learn common features, that fit the task-specific data is challenging.

The required quality metrics to select data from a database are complex and application-specific. Whether some data fit a certain application follows no general rule. Only an expert for the application can define what criteria have to be met and define quality metrics based on the comparison of the available task-specific data and the requested supplementary data. This comparison is based on the unique composition of many different processing methods and metrics for each application scenario. It seems impossible to predefine solutions that fit all scenarios.

The quality of the supplementary data is determined by the suitability to use it jointly with the task-specific data within further processing. Often there is a threshold of sufficient data quality for storing or retrieving it. Furthermore, there are general measures of quality leading to a significant and efficient data collection. These general measures of suitability can be derived from the dataset itself, such as completeness of data and variability of the individual values. Finally, one has to find measures describing the comparability between the supplementary and the task-specific-data as well as the relevance for the use in scientific and industrial purposes. This relevance comparison has to take place both on a high-dimensional and multimodal scale. Related approaches have been developed in the area of literature review in scientific databases [MG20] and the use case of fingerprint identification [BWT18].

We work with spectral data for plant phenotyping for smart agriculture. To get additional similar training data from a spectral database, we have to compare wavelength ranges, recording conditions and the high-dimensional spectra itself.

We will introduce a system with formal description of individual quality metrics of multimodal datasets. The HawkSpex® Featmetric system consists of two parts. On the one hand, there is the description language consisting of a list of operations and parameters that can be customized by the user. On the other hand, there is a serving program that applies these operations on a database.

We propose a business model for using our Featmetric system on databases for data trading. When a customer needs data with individually defined quality metrics, he sends the description to the provider and gets back data with a calculated quality above a given threshold. The system therefore provides a solution for the assessment of potential supplementary data for learning processes.

2 The Featmetric system

The HawkSpex® Featmetric system was first designed to recognize industrial objects by multimodal inherent features without the need of additional hardware labels. For example, the system was used on objects that are too small for QR-codes or other labels as well as on objects with lost or counterfeit labels. The system's underlying concept of comparing reference objects with new objects can easily be used for other applications. Especially, we use it in order to calculate individual quality metrics in data collections. As the system has no special requirements concerning the feature structure, it can even handle data with multimodal features and is not set on specific data formats.

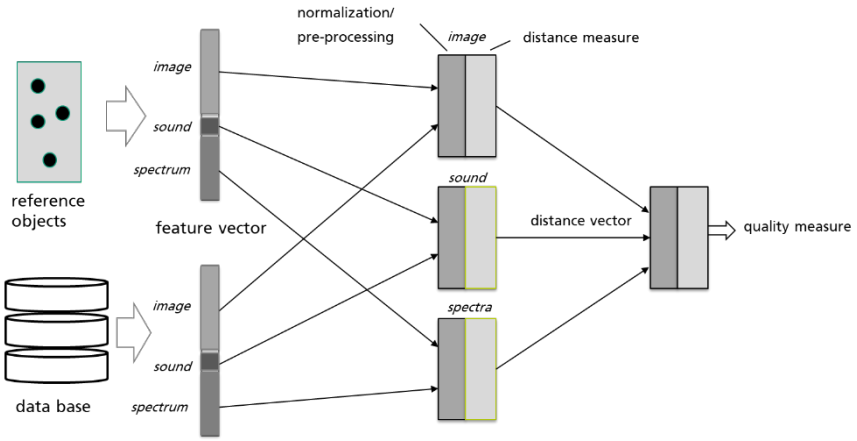


Fig. 1: Typical multistep quality measure with multimodal features.

For the comparison process, the different features have to be normalized individually with a previously selected normalization. The resulting data is then compared to reference data using a distance measurement (see Fig. 1). Therefore, each dataset is converted into a dictionary with key-value pairs. The formal description for the comparison process is a list of operations containing an input key, an output key, the name of the operation function, and if needed, appropriate parameters for these methods. The Featmetric system reads the operations and applies them to all datasets (see Fig. 2). This abstract data-processing description has many advantages. For example, it provides the possibility to easily integrate machine learning models in the process [BHA21]. Additionally, the Featmetric model can easily store, send and compare results and provides the opportunity for versioning.

The integrated operation functions include common statistical methods like different kinds of normalization, comparison, scaling, or collecting. Furthermore, it includes advanced

operations to deal with an unequal number of similar features, such as the size of images. These operations use sorting, random selections, and histogram comparisons. Moreover, it is possible to include the serving of machine learning models in an operation. The machine learning model can be given by *ONNX* or *TensorFlow* descriptions.

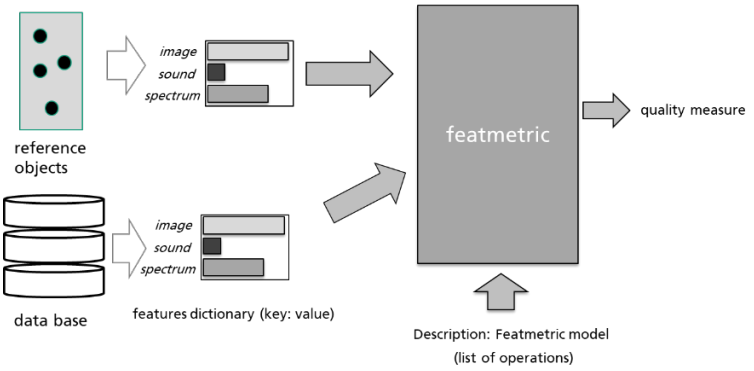


Fig. 2: Featmetric system with abstraction of quality metrics.

The reference object for the comparison can either result from reference datasets, or can be defined by parameters in the operation description. When using reference datasets, the quality metrics may be the application-specific distance measurement of the rated data to the nearest reference.

The parametrization of the integrated operations in the Featmetric-description language was generated such that the predefined parameters fit most use cases. Nevertheless, the system provides the possibility to learn task-specific optimized parameters with the help of neural networks or genetic algorithms applied to some representative examples. This optimization can be easily integrated with one-line operations.

Because most operations can run parallel on all datasets, the Featmetric system provides an effective environment for the data comparison process. Especially for big databases, the program gives the user control over the potential data exchange. Interim results are stored in the datasets during the operation pipeline, but can be erased afterward before sending the results to the customer. Of course, it is also possible to hand them over. Finally, it provides an environment that allows the user to adapt it to its needs. The list of possible operations can be expanded and grown from version to version.

3 Examples of using Featmetric

We tested the Featmetric system with different applications in spectral processing and industrial data analytics.

A typical quality metric for spectral processing is a comparison with a reference spectrum at some specified wavelength. Parts of the spectra are cut out, normalized by L^2, compared by Euclidian distance and the distances are weighted by given scaling factors to get the overall quality measures. Applying these operations with Featmetric on fabric data from a textile production use case helped successfully identifying individual shirts and trousers.

In a different application, we have compared sizes and positions of marked errors in images. The number of features was not fixed. The quality measure included operations to sort by values, compare histograms and different distance metrics. By applying this method to a database of images of circuit boards, we were able to identify one among many boards. Thereby, the user could, for example, identify whether a board is the original or a replacement.

An example of an industrial process is the start-up behavior of an air compressor. This process is characterized by different attributes like the air flow, pressure and temperature over time. Featmetric was used to combine these temporal attributes to characteristic features by approximating Chebyshev polynomials. Based on these features the state and development of different air compressors could be monitored and compared to each other. The resulting information could be used for the machines' maintenance work.

In general, we verified the technical functionality of the Featmetric system and its suitability for identifying individual objects. Furthermore, we could show that our method works across different application areas that are characterized by multi-modal and high-dimensional data. Therefore, the integrated operations proved to be diverse enough to fit the wide range of data hosted for data trading.

4 Business Model for Data Trading

Serving data to customers from big data bases is a successful business model. The importance of data trading is based on the increasing collection of data on the one hand and the improved possibilities to gain knowledge from big data on the other hand.

There is a business relationship between the data provider and the customer. The customer wants to buy the most suitable data-set possible, such that he does not purchase data that has to be discarded later so as to reduce storage space and computing capacity. The data provider wants to sell the data for the highest price possible and often offers the possibility to sell some of the data exclusively. Therefore, giving away less data that is more suitable to the customer gives the provider the possibility to maximize the benefit of the data in total, as well with respect to the monetary value as to the total knowledge that is gained from it.

The Featmetric system is executed on-premise by the data provider and allows him to select a suitable data offer for a given customer and to provide him with descriptive interim results before the transaction. Thereby, he can use his existing hardware sources to provide a more suitable and enriched product to the customer.

The data provider will sell the most suitable data with no arguments about quality. Offering data of reliably high quality will further improve his reputation allowing him to adapt his prices. By selling only the favored few data to the customer the provider has the option of selling this data exclusively, and therefore for a higher price, With the integrated Featmetric system the customer sends his customized Featmetric description to the provider.

In an automated process, the description is checked for errors and tests for consumption of memory and computing time are performed. The test results determine the price offered for using the individual quality metric to find the best data. Thereby, the customer is motivated to optimize the Featmetric description with respect to a small command number and fast operations to get a low price offered. As a result, attacks of endless or highly time consuming and unnecessary calculations can be avoided and the computational capacities will not have unnecessary load.

5 Conclusion

We present an ongoing work to build an extendable description of application-specific quality metrics for data trading. The separation of execution and description has several advantages. The execution in the serving part of Featmetric can be optimized and adapted to the specific database systems while the description part can remain unchanged. The system can be used for the assessment of data for deep learning from large data bases.

There is a close connection to advanced search criteria for data, but the Featmetric system offers more than a database search. With the high number of operations for high-dimensional vector features and multimodal data as well as the possibility to combine them in complex ways, the tool offers more support and is able to meet the various requirements of the customers' unique data requests. Featmetric enhances the possibility to integrate expert knowledge in the search process without giving the customer access to the data before purchasing it.

Last but not least, it may introduce a new business model to high quality data trading. It allows data providers to offer products with an individual pricing based on the customer's definition of quality.

6 References

[BHA21] Backhaus, A.; Herzog, A.; Adler, S.; Jachmann, D.: Deployment architecture for the local delivery of ML-Models to the industrial shop floor. In: Beyerer, J.; Maier, A.; Niggemann, O. (eds) Machine Learning for Cyber Physical Systems. Technologien für die intelligente Automation, vol 13. Springer Vieweg, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-62746-4_4, 2021

[BWT18] Bai, C.; Wang, W.; Zhao, T.; Li, M.: Fast exact fingerprint indexing based on Compact Binary Minutia Cylinder Codes, Neurocomputing, Volume 275, 1711-1724, 2018

[MG20] Marcos-Pablos, S.; García-Peñalvo, F.J.: Information retrieval methodology for aiding scientific database search. *Soft Comput* **24,** 5551–5560 https://doi.org/10.1007/s00500-018-3568-0, 2020

[MBS18] Menz, P.; Backhaus, A.; Seiffert, U.: Transfer Learning for transferring machine-learning based models among various hyperspectral sensors. In: ESANN 2019 proceedings, European Symposium on Artificial Neural Networks, Computional Intelligence and Machine Learning. Bruges (Belgium), 24-26 April 2019, ISBN 978-287-587-065-0. Available from http://www.i6doc.com/en/, 2019

[TSK18] Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C.: A Survey on Deep Transfer Learning. In: Kůrková, V., Manolopoulos, Y., Hammer, B., Iliadis, L., Maglogiannis, I. (eds) Artificial Neural Networks and Machine Learning – ICANN 2018. ICANN 2018. Lecture Notes in Computer Science, vol 11141. Springer, Cham. https://doi.org/10.1007/978-3-030-01424-7_27, 2018

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1023

Approaches for Automated Data Quality Analysis: Syntactic and Semantic Assessment

Agbodzea Pascal Ahiagble [1] and Hannah Stein[2]

Abstract: Data quality significantly influences data usability and plays an important role in data trading. This paper presents a data quality analysis (DQA) of data tables on two levels. The first, the so-called syntactic level, concerns the structure of the elements within the database and the second, the so-called semantic level, concerns the relationship between the elements in the database and the "real world". Based on a literature review the most relevant data quality criteria and corresponding metrics were derived. Subsequently, based on heuristics, a data-centric approach and an unsupervised machine learning clustering algorithm DBSCAN, a service for automated DQA, is designed and implemented (syntactic DQA). In the next step, an automated semantic DQA service as well. The approach is used to examine data tables for example for missing relevant columns (i.e., semantic completeness). A data quality index represents the services' output, which is derived from the automated analysis of various data quality criteria. This enables the assessment of data quality, as well as the detection of potentials for improving quality and thus increasing the value of tradeable data.

Keywords: Data quality assessment, data quality metrics, automated assessment services.

1 Introduction

Data play a very important role for companies' and institutions' activities, e.g., in the management of business processes. High-quality data bring benefits such as increasing customer satisfaction and competitive advantages, guaranteeing higher turnovers for the company [Pip02]. However, data with poor quality can have a negative economic impact on the company [Wan02], [STR96]. Companies and organizations, which work with large amounts of data, require approaches to automatically determine data quality (DQ) across different criteria, as this can be quite a time-consuming process. In data markets and ecosystems, clear responsibilities for DQ assessment are often lacking as well as automatic processes or clear control mechanisms to do so [Cap20], [Lis20]. In addition, data consumers cannot assess DQ before buying the data [Mus12] – a phenomenon that

[1] German Research Center for Artificial Intelligence, Stuhlsatzenhausweg 3, 66123 Saarbruecken, Germany, agbodzea_pascal.ahiagble@dfki.de
[2] Saarland University, Campus A5.4, 66123 Saarbruecken, Germany, hannah.stein@iss.uni-saarland.de

doi:10.18420/inf2022_85

is known as the information paradox [Sta17]. In this work, we present two approaches for decentralized, automatic DQ assessment, appropriate for future application in the context of data ecosystems, i.e., DQ assessment from a syntactic and semantic perspective. The services automatically determine the quality of databases that are generated and used in B2B and B2C contexts. We thereby mainly focus on the quality assessment of databases, defined as a "collection of all stored data and its associated descriptions" [Est00]. The quality of a database is derived from the quality of data or records collected in it.

Next, we present related work and elaborate on the most important DQ criteria for syntactic and semantic DQ assessment. Chapter three shows a methodology approach for developing the automatic assessment services. The technical conceptualization and prototypical implementation are described in chapter four. Chapter five concludes the paper.

2 Related Work

The definition of the quality of a database (DB) is considered from different points of view, e.g., quality is seen as "usability" [Eve07]. In this sense, the determination of quality depends very much on the data consumer and must be contextual [STR96]. The assessment of data quality can be made purely objectively [Lee06], which corresponds to the internal state of the DBs, or subjectively, what exactly the data consumers need or have experienced [Ang12], [Bal98], [Pip02], [STR96]. For this work, we have identified the most important DQ criteria, to be taken into account for semantic and syntactic DQ assessment.

2.1 Data Quality Criteria

The DQ criteria have been intensively researched over the last years and are determined independently of each other. The large number of criteria listed in the literature shows how multifaceted the assessment of data quality can be. Through research and analysis, it is possible to group these criteria, which can vary depending on the context. For example, [STR96] grouped the criteria into four categories, namely intrinsic DQ, accessibility DQ, contextual DQ, and representative DQ. Similarly, [Cai15] have grouped the criteria into availability, usability, reliability, relevance and presentation quality. Criteria such as accuracy, timeliness, reliability, completeness, relevance, accessibility, and interpretability are often used [STR96]. It is significant to know that data type has an impact on the DQ criteria as well as the evaluation and improvement technique [Bat09]. To evaluate DQ by criteria, metrics are used so that they can be represented as a value [Lev95]. In this paper, we focus on the DQ evaluation of structured data to enable automatic assessment service.

2.2 Syntactic Data Quality Assessment

"Syntactic data quality is concerned with the structure of data" [Sha99]. For [STR96] "data quality" is defined as data that are fit for use by data consumers. We have distinguished between evaluable criteria based on the internal elements in the database (objective) and evaluable criteria based on the external elements of the database (subjective). The first group includes all those that are measurable based on the variables provided directly by the database. We provide a summary which is developed based on a literature review in table 1. We name each DQ criterion, give its definition based on literature and then provide an example of how the criterion can be assessed.

Criterion	**Definition**	**Example**
Completeness	"The extent to which data are of sufficient breadth, depth, and scope for the task at hand" [STR96].	Missing values of a DB. Having "NULL" in a field instead of real value.
Integrity	"A measure of the existence, validity, structure, content, and other fundamental properties of the data" [McG08].	Entity integrity, referential integrity, domain integrity, and column integrity reflect only the states of the data [Cod90]
Accuracy	"The closeness between a value v and a value v' that is considered to be the correct representation of the real phenomenon that v is supposed to represent" [Bat06].	In a customer database where a customer has the first name Peter, "v = Peter" is correct and "v = pter" is incorrect.
Timeliness	“The extent to which the age of the data is appropriate for the task at hand" [STR96].	Delay between a change in state of the real world and the resulting change in state of an information system [Bat09], [Wan02].
Consistency	"The data is always presented in the same format and is compatible with the previous data" [Bat06].	Mismatch of datatype (numeric and alphanumeric) as attribute for the same entity.
Free-of-error	In general, this indicates whether "the data are correct" [Lee06].	The extent to which data are correct and reliable. [Pip02]

Table 1: Evaluable criteria based on the internal elements

Regarding the second group of criteria, with the inclusion of "external elements", the fact that the quality depends not only on the internal state of the data, but also on the use and satisfaction that the data bring is considered. Table 2 summarizes these DQ criteria using the same structure as table 1.

Criterion	Definition	Example
Credibility	"The extent to which the data is believed to be true and reliable" [Lee06].	Customers prefer to consume things they believe in without hesitation.
Interpretability and understandability	"The extent to which data are in appropriate languages, symbols, and units, and the definitions are clear" [Pip02].	Customers need to understand the data in order to use it in an appropriate way.
Security	“The extent to which access to the data is appropriately restricted to ensure its safety" [Pip02].	Data protection and confidentiality.
Representativeness	"The extent to which data is represented compactly and in the same format" [Pip02].	"Data are clear, without ambiguity, and easily understood" [Bat06].
Objectivity, Relevance, and Reputation	Objectivity is defined as "the extent to which data are unbiased and unprejudiced", relevance is defined as "the extent to which data are applicable and helpful to the task at hand" and reputation is defined as "the extent to which data are highly regarded in terms of their source or content" [Pip02].	They give data consumers an indication of whether the available datasets meet the baseline for their business processes.

Table 2: Evaluable criteria based on the external elements

2.3 Semantic Data Quality Assessment

Data quality can be defined from the semantic point of view as the "correspondence between the data represented by an information system and the data in the real world" [Ken98]. The criteria for semantic evaluation are similar to those for syntactic evaluation but with different measurement approaches, e.g., completeness, consistency, accuracy, and timeliness [Bat16], [Red98], [Yuv17]. In this context, a database is complete if it contains all the necessary information for the task to be performed [Jar20]. Furthermore, completeness can be defined as the degree to which all relevant attributes of a feature have been encoded [VER99]. Semantic consistency has been defined as the degree to which the evaluated data is free of internal inconsistencies [Ber18]. If we imagine a customer DB of a supermarket, for example, it is not semantically complete if the contact data of customers are not available as an entity. In the case that this characteristic

is present but requests are missing, we speak of syntactic incompleteness.

3 Assessment Approach

The assessment of DQ is specified by metrics. For the syntactic assessment, these metrics can be composed of simple to very complex formulas depending on the variables that come into play. For the construction of the individual metrics, [Pip02] use the simple ratio as a basic method. The quality index (QI) is thereby automatically measured and normalized as a value between 0 (very bad) and 1 (perfect) [Lee06].

According to [Bec08] there are two questions to ask when it comes to the semantic assessment of data quality: First, which real world phenomena should be represented in the information system (i.e., which entities are relevant)? For this purpose, a conceptual model must be created to capture through an ontology: the semantics of a real world. Second, how should relevant phenomena be represented in the information system (i.e., which attributes and relationships are relevant)? Therefore, a conceptual model is compared to the physical data model of an information system.

3.1 Automation of the Quality Assessment of Databases

In order to avoid redundant procedures and save time, it makes more sense to be able to automatically determine the quality of a database, especially now that companies and organizations have to deal with large amounts of data. Companies should be able to keep track of the quality of their databases without much effort, to be able to sell or share parts of them later via data markets and ecosystems. For example, [Bal85] designed a model to detect errors in data-based systems. The provision of data and how the relationships between quality parameters are represented should not be a challenge for companies [Wan02]. Strategies and techniques for assessing and optimizing DQ have been developed in recent years. According to [Bat09], the assessment of DQ is composed of three steps: (1) reconstructing the state of data with the goal of deriving general contextual information, (2) measuring the quality of the database, and (3) searching for improvement strategies and techniques. However, these steps may differ in practice, e.g. Evaluation of the DQ after the measurement or using iterative approaches.

Automation occurs on different levels. Activities that are not performed by humans but by the system can be described as automatic. In DQ assessment, this can be applied in the calculation of the metrics. Appendix of the available data sets, the relevant values for each metric can be collected automatically to perform calculations and return the score as a result. If values cannot be found, a warning of errors is issued instead of the QI as the result. Using heuristic functions [Sch18], [Kri17], targeted, statistical scores can be retrieved from one or more columns in the dataset. Through incremental computation, it is possible to include growing datasets when determining metrics. Machine learning allows, for example, using dataset and column names, to make a prediction about the

types of information present and thus, in case of discrepancies, can list them as errors. Training models with machine learning not only allow to find and fix specific errors such as lower and upper case to fully reflect the quality score [Sch18], [Ruk20], but also allows to independently and continuously monitor the quality level of the DB [Kri17].

4 Proof of Concept

To concretize the points described above, the implementation of a prototype is presented. This prototype is designed to automatically check both the semantic and syntactic quality of databases based on different criteria. According to the automation stages of [She05], our proposal can be classified to stage (2), since the user's intervention is required for the proposed prototype only when providing the records. The architecture of the prototype is illustrated in Figure 1 and has five components. All analysis is done with Pandas[3] and with Danfo.js[4] and visualization is done by Chart.js. As DB, only CSV files are considered here. The evaluation is a decentralized analysis, so it works on edge. The approach used for syntactic evaluation is an adaptation of the approaches proposed in [Pip02], [Sch18]. For consistency, the codes are written in Python.

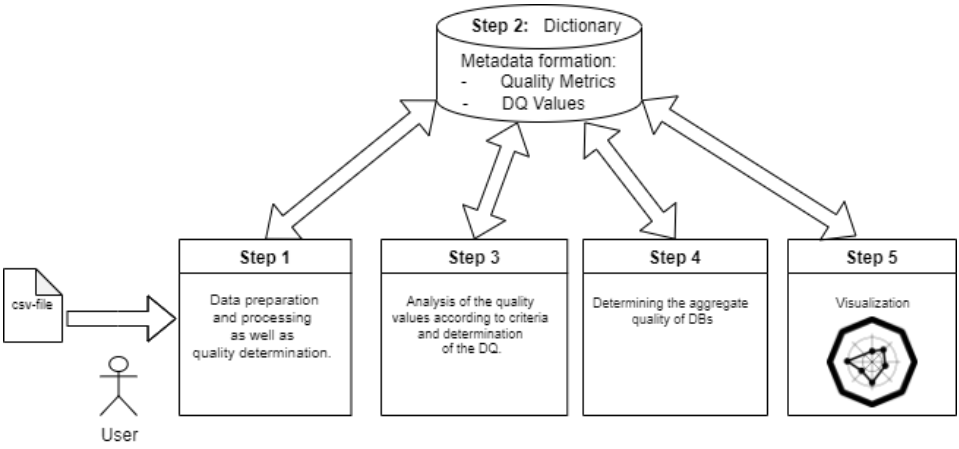


Figure 1: Prototype Architecture[5]

4.1 Preparation of Data Set

As mentioned above, the training data is CSV data. Here, care must be taken regarding

[3] https://pandas.pydata.org/
[4] https://danfo.jsdata.org/
[5] Include for syntactic evaluation step 1, 2, 3 and 4 while including step 1, 2, 3 and 5 for semantic evaluation.

separations (",", "\t", ";", "|"...) in the DBs. The first training data file is a small DB with 17 attributes and 1000 unique values from the platform Kaggle[6]. This dataset is a supermarket data recorded from 3 different shops over a period of 3 months. The second record has 73 attributes and 2686 rows and originates from a manufacturer. After uploading, the records (the CSV files) are stored as metadata in a dictionary[7] in the system, associated with an ID number. The conversion of the CSV file is done by using Pandas. First, we read the file as follows:

```
csv = pd.read_csv(csv_file, sep=delimiter)
```

Then we store the tables in the dictionary in the system to be able to continue working with them, as the code line shows:

```
data = csv.to.dict
```

4.2 Assessment of Quality Values

Pandas is used in different ways depending on the criterion to perform the evaluation. For syntactic quality assessment, four criteria are specifically included, namely completeness, consistency, accuracy and integrity of IDs [Ehr19]. This represents a data-centric approach. Additional inputs are required to consider other criteria [Lee06]. For Semantic Quality Assessment, completeness and consistency are also considered as criteria here [Ber18].

Syntactic Completeness

Pandas examines the dataset for empty entries, also called "Null Values". The number of "Null Values" is incrementally enumerated and then divided by the size of the dataset. Finally, the result of the division is normalized according to the "Simple Ratio" formula. The whole process looks like the following code in Python:

```
null_counter = int(csv[col].isnull().sum())
incomplete_cnt += val['null_counter']
metrics['completeness'] = 1-(incomplete_cnt/csv.size)
```

Syntactic Consistency and Accuracy

Here, mainly data type inconsistencies are considered. To find inconsistent values and inaccuracies, a heuristic function and a data-centric approach are used. As assumption, our method distinguishes between two types of inconsistencies, namely the so-called

[6] https://www.kaggle.com/aungpyaeap/supermarket-sales, accessed : 12.02.2022.
[7] https://realpython.com/python-dicts/. accessed 02/2021.

data type_inconsistency and the outliers per expression [Bat06]. The first type of inconsistency is composed of alternations of types within the same column. For accuracy, this is taken as the only violation characteristic. The second type, the outliers can be data entry errors, valid data but from a different population, or very rarely occurring data from the same population. As a basis of analysis, a heuristic based on statistical techniques is used: (1) The majority of values per expression are correct, with inconsistent values occurring infrequently and (2) errors can be different but equally distributed, or they can be the same but represent little compared to the correct entries. In order for this heuristic to be applied, a distinction was made between continuous and categorical data by inference. For the clustering of the data type DBSCAN[8] is used.

Semantic Completeness

For this category of evaluation, the absence of certain groups of attributes are considered as a violation [Jar20]. To overcome the challenge of a standard of minimum attributes that must be present in a database, database models of CRM vendors were analyzed and some interviews with different vendors were conducted. A database in the B2B domain should contain at least demographics, firmographics, technological, chronographic quantitative and qualitative data[9]. In this work master data in the B2C area were assumed as name, gender, address, email address and city. Within the database, these attributes are searched. In order to not miss any information, we first search for the synonyms of the attributes, which we include in the further processing as follows:

```
mydic = dict()
for i in basedata[10]:
    for val in wornet.synsets(i):
          for j in val.lemmas():
                array_synonyms.append(i.name())
```

Based on the number of master data and the number of matching attributes found in the DB, the semantic completeness is determined as follows:

```
Size_base = len(basedata)
semCompleteness = count[11] / size_base
```

The completeness decreases depending on how the attributes are available.

Semantic Consistency

The consistency on location, email address, gender and age are checked here. When

[8] https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html, accessed 04/07/2022.
[9] https://www.cognism.com/b2b-data, accessed 04.07.2022.
[10] The master data collection
[11] Number of attributes occurring

checking the location data, the aim is to ensure that the specified location in the database actually exists in the real world. Therefore, individual locations are checked for existence in real-time using the free geographic database platform GeoNames API[12]. The code section for this looks like this:

```
Check_loc=requests.requests("GET",f"https://www.geona
                mes.org/search.html?q={Ci[13]}&country=")
```

Email address consistency was realized using the Python Library email-validator[14].

```
Regex = '^[a-z0-9]+[\._]?[a-z0-9]+[@]\w+[.]\w{2,3}$'
    If True == validate_email(m[15]) or
                (re.search(regex,m) == True):
                 isvalid.append(m)
```

For gender information, two gender types were considered as the base gender. The two genders are set in different formats and in three languages, namely English, German and French as default and searched in the database. We summarise it as code as follows:

```
Standard_Gender = ['female', 'male', 'weiblich',
          'männlich','feminin','masculin','f', 'm','w']
```

For age, all data above 100 and below 0 are considered as outliers as shown below:

```
If age < 100 and age > 0:
            not_outliers.append(age)
```

These four criteria are adjusted for the Quality Index and Simple Ratio.

4.3 Visualization

The ratings are presented as an index. The algorithm receives all uploaded data sets with their matching criteria as well as metrics as input. The whole is presented in the form of a radar, where the radius is equal to 1, which in turn corresponds to the QI of the "Simple Ratio = [0, 1]". Figure 2 shows how the visualization is presented

[12] http://www.geonames.org/, accessed : 05.07.2022.
[13] Current city
[14] https://pypi.org/project/email-validator/, accessed : 05.07.2021
[15] Current email address

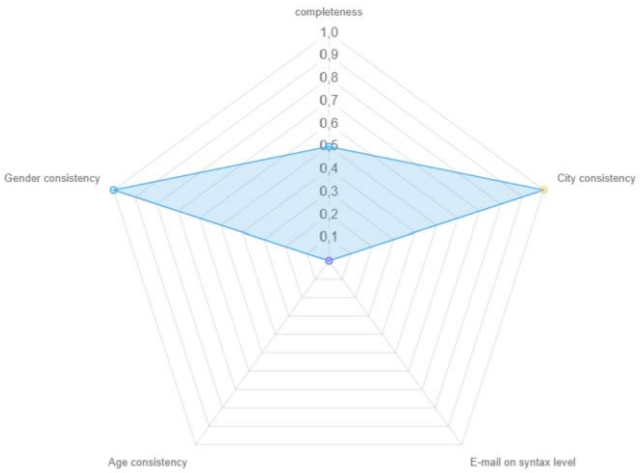


Figure 2: Visualisation of the supermarkt.csv evaluation

On the graph we can see that City Consistency and Gender Consistency have a QI equal to 1 (very good), which means that the data in the DB matches the data in the real world. So the cities really exist and the genders match the given standard. Age and e-mail consistency are equal to 0 in this case because they were not to be checked in the DB (see point: 3 description Fig.4). This leads to the QI of completeness being equal to 0.5 because this DB does not have a "Name" attribute in addition to these two attributes (see point: 1 description Fig.4).

4.4 Result

The results of the Proof of Concept (PoC) look in our interface as shown in the figures provided below. For the syntactic evaluation, the QI of the dataset and the aggregated QI are displayed first. For each dataset, the criteria values are used to determine how the QI was obtained. For the semantic evaluation, the QI per criteria is determined on the left side of the interface and, if necessary, the appropriate improvement options are indicated on the right side.

Record	Quality result	Data preview
Real life database.csv	88.20	Details
Average quality	**88.20**	

The quality criteria

Criteria	Values
Accuracy	0.9996327991921582
Consistency	0.9996327991921582
Completeness	0.5417843919256622
Id_integrity	0.9868215710074563

Figure 3: Syntactic evaluation of Real_life_database.csv

- The semantic completeness is: 0.5
- The semantic consistency in terms of locality is: 1.0
- The quality index of the e-mail at the syntax level is: 0.0
- The quality index of e-mail at the semantic level is: 0.0
- The quality index of gender at the semantic level is: 1.0
- The quality index of age at the semantic level is: 0.0

1. For semantic completeness, we recommend that you update your database with the following attributes:
['Name', 'Address', 'E-mail address']
2. The following given places are not to be found:
[]
3. For semantic consistency, we recommend that you check and update the completed email address:
['No email address was to be verified']
4. For semantic consistency, we recommend that you check and update the completed email address:
['No email address was to be verified']

Figure 4: Semantic evaluation of supermarkt.csv

5 Conclusion

Our work comes not without limitations which need to be overcome by future work. First, only four criteria were considered for the syntactic evaluation. Criteria like timeliness and accessibility, which are important for the companies, were left out. The inclusion of such criteria requires temporal information. Assuming that this information is provided by an external system, machine learning can be used in further assessment areas to extend the proposed model to include these criteria. For semantic evaluation, there is currently very little source to be found in the literature. Our PoC focuses on the DQ assessment of customer databases and is based on the analysis of customer database providers[16] [17] [18]. Further research in the context of semantic DQ assessment needs to consider further domains and types of data. In addition, we worked exclusively with structured data. Semi-structured data such as JSON files as well as unstructured data such as text or conversations, on the other hand, are difficult to evaluate without first adapting them [Blu03], [Tei16], [Ehr19]. An automatic procedure for evaluating this type of data would be important in future work.

In this work, we developed a prototype consisting of two services: one for syntactic and one for semantic DQ assessment. For the PoC, an open source database from the platform Kaggle[19] and a "real life" database from a manufacturer were used. Especially the semantic approach shows opportunities for more efficient and transparent DQ assessment to be applied for data sharing in data ecosystems.

6 Acknowledgement

This work is part-funded by the research project "Future Data Assets" (grant number: 01MD19010C) funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) within the "Smart Data Economy" technology program, managed by the DLR project management agency.

References

[Ang12] Angeles, M. d. P.; García-Ugalde, F. J.: Subjective Assessment of Data Quality considering their Interdependencies and Relevance according to the Type of Information Systems, International Journal on Advances in Software, Bd. 5, Nr. 3&4, pp. 389-400,

[16] https://www.oracle.com/webfolder/s/quicktours/cx/pt-customer-data-mgmt/index.html, accessed: 11.2021.
[17] https://help.salesforce.com/s/, accessed: 11.2021.
[18] https://www.teamleader.eu/blog/customer-data-management-how-to, accessed: 01.2022.
[19] https://www.kaggle.com/aungpyaeap/supermarket-sales, accessed : 12.02.2022.

2012.

[Bal85] Ballou, D.; Pazer, H.: Modeling Data and Process Quality in Multi-Input, Multi-Output Information Systems, Management Science, Bd. 31, Nr. 2, pp. 150-162, 1985.

[Bal98] Ballou, D.; Wang, R.; Pazer, H.; Tayi, G. K.: Modeling Information Manufacturing Systems to Determine Information, Management Science, Bd. 4, Nr. 44, pp. 462-484, 1998.

[Bat06] Batini, C. ; Scannapieca, M.: Data Quality Concepts, Methodologies and Techniques, Springer-Verlag Berlin Heidelberg, 2006.

[Bat09] Batini, C.; Cappiello, C.; Francalanci, C.; Maurino, A.: Methodologies for Data Quality Assessment and Improvement, ACM Computing Surveys, Bd. 41, Nr. 3, pp. 16:1-16:49, 2009.

[Bat16] Batini, C.; Scannapieca, M.: Data Quality Dimensions, Data and Information Quality, pp. 11-51, 2016.

[Bec08] Becker, J.; Matzner, M.; Mueller, O.; Winkelmann, A.: Towards a Semantic Data Quality Management using Ontologies to Assess Master Data Quality in Retailing, Proceedings of the Fourteenth Americas Conference on Information Systems, pp. 1-11, 14-17 August 2008.

[Ber18] Bernd, H.; Klier, M. Schiller, A.; Wagner, G.: Assessing Data Quality – A Probability-based Metric for Semantic Consistency.," Decision Support Systems, pp. 95-106, 2018.

[Blu03] Blumberg, R.; Atre, S.; The Problem with Unstructured Data, DM Review, 2003.

[Cai15] Cai, L. ; Zhu, Y.: The Challenges of Data Quality and Data Quality and Data Quality Assessment in the Big Data Era, Data Science Journal, Bd. 14, Nr. 2, pp. 1-10, 2015.

[Cap20] Capiello, C.; Gal, A.; Jarke M.; Rehof, J.: Data Ecosystems: Sovereign Data Exchange among Organizations (Dagstuhl Seminar 19391), Report from Dagstuhl Seminar 19391, pp. 66 - 134, 2020.

[Cod90] Codd, E. F.: The Relatinal Model for Database Management, Eddison-Wesley, 1990.

[Ehr19] Ehrlinger, L.; Rusz, E.; Wöß, W.: A Survey of Data Quality Measurment and Monitoring Tools, 2019.

[Est00] Ester, M.; Sander, J.: Knowledge Discovery in Databases, Springer-Verlag Berlin Heidelberg, 2000.

[Eve07] Even, A.; Shankaranarayanan, G.: Utility-Driven Assessment of Data QUality, The DATA BASE for Advances in Information Systems, Bd. 32, Nr. 2, pp. 75-91, 2007.

[Jar20] Jarwar, M.,C.; Chong, I.: Web Objects Based Contextual Data Quality Assessment Model for Semantic Data Application, Appl. Sci. 2190(10), pp. 1-33, 2020.

[Ken98] Orr, K.: Data quality and systems theory. Communications of the ACM, 41, pp. 66 – 71, 1998.

[Kri17] Krishnan, S.; Franklin, M. J.; Goldberg, K.; Wu, E.; BoostClean: Automated Error Detection and Repair for Machine learning, 2017.

[Lee06] Lee, Y. W. ; Pipino,L. L. ; Funk , J. D. ; Wang , R. Y. : Journey to Data Quality, The MIT Press, 2006.

[Lev95] Levitin, A.; Redman, T.; The notion of data and its quality demension, Information processing & Managemant, Bd. 30, Nr. 3, pp. 9-19, 1995.

[Lis20] Lis, D.; Otto, B.: Data Governance in Data Ecosystems – Insights from Organizations, AMCIS 2020 Proceedings, pp. 1- 10, 2020.

[McG08] McGilvray, D.; Executing Data Quality Projects: TEN STEPS to Quality Data and Trusted Information, Morgan Kaufmann, 2008.

[Mus12] Muschalle, A.; Stahl, F.; Löser, A.; Gottfried, V.: Pricing approaches for data markets., International workshop on business intelligence for the real-time enterprise, pp. 129–144, 2012.

[Pip02] Pipino, L. L.; Lee, Y. W.; Wang, R. Y. ; Data Quality Assessment, COMMUNICATIONS OF THE ACM, Bd. 45, Nr. 4, pp. 211-218, 2002.

[Red98] Redman, C.: The Impact of Poor Data Quality on theTypical Enterprise, COMMUNICATIONS OF THE ACM, Bd. 2, pp. 78 - 82, Februar 1998.

[Ruk20] Rukat, T.; Lange, D.; Schelter, S. ; Biessmann, F.: Towards Automated Data Quality Management for Machine Learning, Amazon Science, 2020.

[Sch18] Schelter, S.; Lange, D.; Schmidt, P.; Celikel, M.; Biessmann, F.; Grafberger, A.: Automating Large-Scale Data Quality Verification, Proceedings of the VLDB Endowment, Bd. 11, Nr. 12, pp. 1781-1792, 2018.

[Sha99] Shanks, G.; Corbitt, B.: Understanding Data Quality: Social and Cultural Aspects, Proc. 10 785 th Australasian Conference on Information Systems, 1999, pp. 785-797, 1999.

[She05] Sheridan T. B.; Parasuraman, R.: Human-Automation Interaction, Reviews of Human Factors and Ergonomics, Bd. 1, Nr. 1, pp. 89-129, 2005.

[Sta17] Stahl, F.; Schomm, F.; Vomfell, L.; Vossen, G.: Marketplaces for Digital Data: Quo Vadis?, Computer and Information Science, Bd. 4, Nr. 10, pp. 22-37, 2017.

[STR96] Wang, R. Y., & Strong, D. M.: Beyond Accuracy: What Data Quality Means to Data Consumers, Journal of Management Information System, 12(4), pp. 5-34, 1996.

[Tei16] Teinemaa, I.; Maggi, M.; Fracescomarino, C. D.: Predictive Business proces Monitoring with Structured and Unstructured Data, Business Process Management, Bd. 14, pp. 402-417, 2016.

[VER99] VEREGIN, H.: Data quality parameters, Geographical information systems, pp. 177 - 189, 1999.

[Wan02] Wand, Y. ; Wang , R. Y.: Anchoring data quality dimensions in ontological foundations. Communications of the ACM, 39(11), pp. 86-95, 1996.

[Yuv17] Yuval Z.; Even, A.:Development and evaluation of a continuous-time Markov chain model for detecting and handling data currency declines, Decision Support Systems, pp. 82 - 93, Januar 2017.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1037

TOWARDS A DATA QUALITY INDEX FOR DATA VALUATION IN THE DATA ECONOMY

Dusan Dokic [1], Hannah Stein[2]

Abstract: Data represent a key resource for firm success, being used for strategic decision making and increasing business process efficiency. Despite the large potential of data sharing within data ecosystems or markets, firms are reluctant to do this, due to fear of losing competitive edges, lack of trust and ambiguity regarding data value. According to prior research, data value vastly depends on usage and quality. This paper focuses on data quality, as the lack of methods for quantifying data quality is one main reason for missing comprehensible data valuation approaches. We analyze 15 existing data quality indices (DQI) from theory and practice, identify relevant data quality dimensions and discuss metrics for applicability in data valuation approaches for data ecosystems and markets. Based on a quantitative study, we propose a DQI concept for developing transparent, objective data valuation methods, while providing a better understanding of inter- and intra-organizational data value.

Keywords: data quality index; data quality metrics; data economy

1 Introduction

The importance of data for the success of modern companies is non-debatable as data-driven business models can be identified at the heart of multi-billion dollar companies like Facebook or Telefonica [Fa20]. Data-driven business models are implemented to such an extent in industries like insurance, banking, telecommunication, and tourism, that some already regard them as data monetization industries [HN20]. One example for such business model is the extraction and selling of raw behavioural customer data to third parties [WS22]. The relevance of data-driven business models is expected to increase, as data sharing marketplaces and ecosystems are being increasingly addressed by current research and practitioners [Ab21]. Data sharing within data ecosystems and the resulting increase in data utilization is expected to create up to 30 percent of the world's gross domestic product [JT18] and to increase involved companies' revenue. Engagement in data ecosystems is regarded as necessity for future economic survival of organizations [SHS13]. This increases the pressure on management to leverage existing technologies like artificial intelligence and machine learning, which highly depend on data of high quality [ACS16]. A key challenge in developing and operating such data ecosystems or sharing platforms is the determination of data value and to express it in monetary units, so that participating data providers can be compensated accordingly [Ge21].

[1] German Research Center for Artificial Intelligence, Stuhlsatzenhausweg 3, 66123 Saarbruecken, Germany, dusan.dokic@dfki.de
[2] Saarland University, Campus A5.4, 66123 Saarbruecken, Germany, hannah.stein@iss.uni-saarland.de

doi:10.18420/inf2022_86

However, to the best of the authors knowledge, replicable and universally applicable methods to determine monetary data value seem to be missing [ST17]. While many factors are influencing the value of data [MW99], related work points to the fact that data quality (DQ) represents a main value driver for data [OW08, SM22]. This means that the value of data cannot be assessed without assessing its quality first. Since data are the main resource or 'input' in data markets, DQ valuation methods are likely to be at the heart of any business model of such platforms and ecosystems, otherwise the value proposition will not be delivered. Literature is unanimous that DQ represents a multidimensional concept [WS96, PLW02]. Therefore, this paper will focus on the individual dimensions of DQ. To measure the quality of data, it is necessary to measure the single dimensions of it first. Research on data quality dimensions (DQD), their definitions and metrics already exists, and will be discussed in more detail later on in the paper.

The aim of this work is to collect existing attempts of creating a data quality index (DQI) and to analyze them based on the DQD and metrics they use. Based on these results, we propose a first DQI model that that is equipped with guidelines about the structure of such a DQI, what DQD are considered and what metrics are used. Our proposed DQI model represents a sturdy starting point for assessing the quality of data in data ecosystems. The remainder of this paper is structured as follows: in the second chapter, related work in regards to data quality and data ecosystems is being discussed. This includes an overview of 15 current examples of DQI. In the third chapter of this paper, the study design is laid out and the study results are analysed for possible implications regarding our own DQI model. Afterwards, the discovered data quality metrics are compared and filtered for their applicability on data ecosystems in chapter 4. The fifth chapter synthesizes the results of the previous chapters into a first proposal for a DQI for data ecosystems. Finally, a summary of the paper results is given and limitations and future work are discussed.

2 Data Quality for Data Ecosystems

This paper focuses on analysis and differentiation of DQD and their metrics, since these constitute the majority of a data quality index. We will elaborate on the metrics of the single data quality dimensions in a separate chapter, as the metrics make up the most important part for the data quality index.

Literature suggests that information might be more valuable than the actual technology of a company and the discussion on measuring the value of information in a company goes even back to the 1990's [Gl93]. Numerous (research) domains such as mathematics, accounting and logistics developed specific approaches in order to measure data and information value, as well as DQ [Bo14, Wi12]. Other works that focus on the importance of DQ in general and justify developing a measuring method for the same can be found across literature [NS17, ORH05, Re19].

Data Quality Dimension	Definition
Accessibility	The extent to which data are available or easily and quickly retrievable.
Appropriate amount of data	The extent to which the volume of data is appropriate for the task at hand.
Believability	The extent to which data are accepted or regarded as true, real, and credible.
Completeness	The extent to which data are of sufficient breadth, depth, and scope for the task at hand.
Concise Representation	The extent to which data are compactly represented.
Consistent Representation	The extent to which data are presented in the same format.
Value-added	The extent to which data are beneficial and provide advantages from their use.
Accuracy	The extent to which data are correct and reliable.
Interpretability	The extent to which data are in appropriate languages, symbols, and units and the definitions are clear.
Objectivity	The extent to which data are unbiased, unprejudiced, and impartial.
Relevancy	The extent to which data are applicable and helpful for the task at hand.
Reputation	The extent to which data are highly regarded in terms of its source or content.
Security	The extent to which access to data are restricted appropriately to maintain its security.
Timeliness	The extent to which data are sufficiently up-to-date for the task at hand.
Understandability	The extent to which data are easily comprehended.
Portability	Transferability to other usage targets, e.g., sales data, are additionally used for product recommendations.

Tab. 1: Definitions of the DQD according to [PLW02]

DQ is often defined as 'the fitness for use' and is regarded as a multi-dimensional concept [WS96]. Thereby, DQ defines and provides DQD that are being reused throughout literature. For example, a conceptual framework of DQ was developed, including 16 dimensions and group them into four categories: intrinsic, contextual, representational, and accessibility [WS96]. In addition, these authors developed further DQ assessment methods and metrics and address data from a management perspective [KSW02, Le02, PLW02, WML01]. A utility-driven measurement framework for DQD within a specific usage context points to the fact that high DQ has a positive effect on business value[ES07]. It ties the value of data to specific assets in the company and links these to DQD like 'Completeness' or 'Accuracy'. Furthermore, practical examples for the use of these dimension in frameworks and methods for assessing DQ are given. All above-mentioned works are crucial as a basis for the development of a holistic approach for measuring DQ within the context of data valuation in data ecosystems using a DQI. For the development of our own DQI we will use the shortened list of [PLW02] as the basis for developing the index and the associated questionnaire. It includes the dimensions that are being regarded as relevant in the broad literature and for which pre-evaluated questionnaire items already exist [WS96]. Still, not

all dimensions are considered as equally important in literature. Some dimensions are being renamed while keeping the exact definition. For example, 'Completeness', 'Timeliness', and 'Accuracy' are being mentioned more frequently than others [CZ15],[Se16] [Ta16]. Meanwhile, dimensions like 'Believability' are sometimes exchanged with terms like 'Credibility' [CZ15]. The full list of the DQD and their definitions are given in Table 1.

2.1 Data Quality Indices

An index is a model that is being used to capture the information of a multi-dimensional theoretical construct by aggregating the different variables that make up the construct into a single number [DB16]. An index turns highly complex constructs into an understandable and comparable measure. This reduction in complexity comes with a loss of information as the involved indicators are measuring different aspects of the construct that can't be directly compared with each other. While this trade-off between complexity and loss of information is regarded as the main argument against such an index, there are plenty of examples across various disciplines that leverage such indices to convey information in a simple and understandable way to third parties [Un22]. Following the definition of an index provided above, DQ fulfils all the requirements for the use of an index: it is complex in its structure and depends on various dimensions that all measure a different aspect of it. Literature suggested to develop and apply DQI more widely [PLW02]. Since then, a modest number of undertakings can be identified throughout the literature and practice that have attempted to develop a DQI. In the following, we give an overview of these DQIs and analyse them based on their weaknesses and strengths, their comparability, as well as the DQD and their corresponding metrics. Finally, we extract implications for the creation of our own concept of a DQI.

Attempts to create an index for the assessment of DQ are scarce in literature, and those that can be found are mostly context-specific. Table 2 gives an overview of some of these indices. Further Indices can be found in Table 4 and 5 in the appendix. In total, we were able to identify 15 DQIs from literature and practice. For instance, a DQI was used to identify Nursing Quality Indicators for Reporting and Evaluation, using the DQD 'Completeness', 'Timeliness', 'Accuracy', and 'Consistency'[Na20]. They use a data set about prevention of falls and fall injuries to show that an assessment through a DQI was possible and that a specific value could be attached to the data set at hand. In another example, a DQI is used to establish a data cockpit for the Bayer CropScience AG [EHO11]. They aim to harmonize the data infrastructure of the 120 different national subsidiaries of the Bayer CropScience AG by combining the master data of all subsidiaries into a so-called 'golden box' and distributing it from there to the regional systems. Their DQI is kept simple based on a ratio of a faulty data set to the number of data sets in total. Inaccurate data sets are defined as the data sets that violate one of the predefined business rules assigned to validation groups. As a final example for the current use of a DQI, the European Central Bank (ECB) approach in Frankfurt can be identified [SB16]. This DQI is aimed at banks that the ECB is supervising and uses the

dimensions 'Punctuality', 'Accuracy', 'Consistency', and 'Completeness' amongst others. This DQI is supposed to ensure the information submitted by the banks to the ECB is of sufficient quality and to enable a comparison of the organizations.

Origin	Data Type	Use Case	DQD	Strength	Weakness
[EHO11]	Process	DQ cockpit for improving, monitoring, comparing master data across subsidiaries ERP systems.	Unspecified; Business Rules	Comparable measure for corporate DQ monitoring; Measures process relevant data flaws	Company-specific; aggregation on validation group level instead of DQD.
[EBH13]	Customer	DQI-based system for controlling DQ.	Completeness, Accuracy	Example how DQI can improve DQ over time.	Ambiguity how DQD are assessed, measured and corrected.
[SB16]	Finance	ECB-Guidelines for DQ assessment of data from supervised banks.	Punctuality, Completeness, Accuracy, Plausibility	Traffic light system based on rating from 1 (good) to 4 (very bad / missing).	No specific metrics provided; weights unexplained
[Le18]	Corporate	Automatic discovering and review of business rules, incl. a DQI.	Referential integrity, Date, Code, Encryption and Numbering Pattern	Automatic review and discovery of business rules reduces complexity.	Indicators and metrics lack comprehensibility; DQI score is not explained.
[Na20]	Healthcare	DQI assesses DQ for nursing quality, evaluation and reporting	Completeness, Consistency, Accuracy, Timing	DQD and metrics are explained including formulas.	Chosen DQD are non-holistic; context specific weights; partial subjectivity.
[SWL20]	Corporate	DQI-based DQ assessment system for a cross-company batch management system in the tobacco industry.	Integrity, Accuracy, Timeliness, Uniqueness, Consistency, Validity, Stability	Example for the use of a DQI for comparing data between enterprises in the same industry.	Metrics for DQD values are unclear as the values are already 'provided'.

Tab. 2: Overview of the existing DQIs from literature and practice - part 1

In order to be able to determine which DQD should be included in our model, a quantitative study was conducted. The design and execution as well as the results of this study will be discussed in detail in the next section.

3 Study Design

We developed a quantitative survey aiming to rank the individual DQD according to their importance for the companies and to determine possible weights for the later index. The questionnaire was designed to address mainly data professionals such as Chief Data Officer of companies. In the following, an overview of the questionnaire's development as well as a description of the conducted study is given.

3.1 Development of the study questionnaire

We follow the guidelines for creating a standardized questionnaire and which aspects have to be considered [DB16]. As recommended by [DB16], the individual questionnaire items were taken from existing literature [WS96]. A first draft of the questionnaire was tested and discussed in a questionnaire expert conference where a total of 4 people, experienced in the development of questionnaires, were interviewed and asked for comprehensive feedback. The feedback was considered and incorporated until no more suggestions for improvement were made.

The questionnaire is structured as follows. In the beginning, a brief instruction is given on the questions' structure and what the single scale points of the 9-point-Likert scale mean. Thereby we aimed to simplify handling the questionnaire for the participants and to prevent ambiguities. The structure of the questionnaire consists of title, instructions on how the questions should be answered, content-related question blocks, statistical information, a feedback field, and a concluding section. The first questionnaire item asks the participants' perception of the importance of the three data value dimensions, 'Quality,' 'Usage,' 'Timeliness', concerning the value of data. These three dimensions have been identified as potentially important for the valuation of data in a qualitative study [SM22] and the survey at hand aims to verify this assumption. In the questionnaire, the associated definitions of the dimension are given to clarify what is meant by the used terms. The central part of the questionnaire aims to measure the perceived importance regarding different DQD in terms of DQ. For this purpose, the list of already identified data quality dimensions by [PLW02] was adopted for the study. The list contains a total of 16 quality dimensions listed in Table 1. It should be noted that the name of the dimension 'Ease of Manipulation' was changed to 'Portability'. It should also be mentioned here that the dimension 'Free-of-Error' named by [PLW02] is also often titled 'Accuracy' in the literature. No original questionnaire items were developed for the study, instead already validated items, rating scales and ranking methods were adopted [WS96]. The original rating scale from 1 to 9 used by [WS96] is used: 1 stands for 'Extremely important', 5 for 'Important', and 9 for 'Not important at all'. For clarity and streamlining purposes, the 16 DQDs are divided into four blocks of 4 DQDs each. In each block, the question is repeated, and the definitions of the dimensions are given. In the last part of the questionnaire, information on the participant's company is being captured [DB16]. This includes information on the domain, number of employees and

target customer group. Furthermore, information on the participant's years of experience working with data and their current job title are queried.

The survey was conducted via Unipark, an online survey tool provider. The target group were companies operating in the business-to-business sector. We followed numerous acquisition strategies, e.g., through social business networks, our own network, cold calls, etc. In sum, 31 valid questionnaires could be used for analysis. The analysis has shown that 32 percent of the participants are working in companies with less than 50 employees, while 42 percent are working in companies that have employees in the range of 50 to 499 employees and 26 percent of the participants have answered that they are employed in a company with 500 or more employees. 58 percent of these 31 participants declared that they have 5 or fewer years of experience in working with data. The remaining 42 percent declared that they have more than 6 years of experience.

3.2 Study results

The responses were aggregated, and a mean was taken for each of the DQD rated by the subjects (definitions of DQD can be found in Table 1). The dimensions were then ranked in descending order of importance (lowest to highest mean value). The analysis results are given in Table 2 and show that the dimension 'Believability' has the lowest mean value (2.19) and is perceived as the most important by the participants. The upper quarter of the table also includes the dimensions 'Value-added' (2.26), 'Free-of-Error' or 'Accuracy' (2.52), and 'Timeliness' (2.61). In the last quarter are the dimensions 'Reputation' (3.29), 'Understandability' (3.65), 'Concise Representation' (3.81), and 'Portability' (4.19).

Data Quality Dimension	Indicator	Mean
Believability	β_1	2.19
Value-added	β_2	2.26
Accuracy	β_3	2.52
Timeliness	β_4	2.61
Interpretability	β_5	2.65
Relevancy	β_6	2.65
Consistent Representation	β_7	2.87
Accessability	β_8	3.00
Objectivity	β_9	3.06
Security	β_{10}	3.16
Completeness	β_{11}	3.19
Appropriate Amount of Data	β_{12}	3.26
Reputation	β_{13}	3.39
Understandability	β_{14}	3.65
Concise Representation	β_{15}	3,81
Portability	β_{16}	4.19

Tab. 3: Importance of DQD by ascending means

4 Data Quality Metrics

In the following, the different approaches to metrics for the previously described and evaluated DQD are presented and explained. [PLW02] gives three forms that can be used to develop such metrics and provide suggestions on which to use for what dimension. The simple ratio of undesirable outcomes to total outcomes subtracted from 1 can be used for dimensions like 'Accuracy', 'Completeness', 'Concise representation', and 'Relevancy'. Applying this to the dimension of 'Consistency' leads to the following metrics:

Consistency

$$= 1 - \frac{number\ of\ violations\ of\ a\ specific\ consistency\ type}{total\ number\ of\ consistency\ checks} \tag{1}$$

Accuracy

$$= 1 - \frac{Number\ of\ incorrect\ data\ sets}{Total\ number\ of\ data\ units} \tag{2}$$

To be able to use these metrics, a definition of what makes data 'incorrect' has to be introduced beforehand. Business rules can then be used to count the number of data sets that do not adhere to the 'Accuracy'-rule.

Completeness

$$= 1 - \frac{Number\ of\ data\ sets\ with\ missing\ values}{Total\ number\ of\ data\ sets} \tag{3}$$

For metrics like 'Appropriate amount of data', 'Accessibility' or 'Timeliness', a min-max operator is suggested [Le06]. It can handle multi-dimensional input and leaves the option to either be liberal or conservative in the aggregation of the values of the individual indicators. While the authors do not give specific metrics for the dimensions they name, their work is used as the theoretical basis for future studies that provide them [Le06].

Appropriate amount of data

$$= min[\frac{Number\ of\ data\ sets\ provided}{Number\ of\ data\ sets\ needed}, \frac{Number\ of\ data\ sets\ needed}{Number\ of\ data\ sets\ provided}] \tag{4}$$

For this metric, the amount of data has to be specified beforehand. The min operator can then be used to choose the smallest value of the given ratios.

Timeliness

Metric according to [Le06]

$$= \{max[(1 - \frac{Currency}{Volatility}, 0]\}^S \tag{5}$$

with **Currency**

$$= (delivery\ time - input\ time) + age \tag{6}$$

and **Volatility**

$$= length\ of\ time\ over\ which\ the\ data\ remains\ valid \tag{7}$$

Delivery time stands for the time at which the data was delivered to the user, while input time means the time at which the system received the data. Age represents the age at which the system received the data. S denotes the exponent value and parameter that controls the sensitivity of the metrics. If a value of S = 1 stands for no adjustment, a value of S < 1 stands for less sensitivity and makes the metric for timeliness bigger. Following the same logic, a value of S > 1 makes the metric more sensitive and returns a smaller metric value[Le06, Ba98].

Metric according to [Ge18]

$$= 100 * (\frac{Number\ of\ data\ sets\ that\ have\ been\ updated\ since\ a\ chosen\ date}{Total\ Number\ of\ data\ sets}) \tag{8}$$

Accessibility

$$= \{max[(1 - \frac{Request\ time}{Relevance\ time}, 0]\}^S \tag{9}$$

In this metric, 'Request time' represents the current interval of time (in seconds) that it takes from the request by the user to the deliver of the data to the user. 'Relevance time' means the interval of time from request by the user to the time at which the data is no longer of any use to the user. This metric is based around the definition that the earlier data is delivered, the more valuable it is. As the first term would eventually become negative due to the long delivery time, the max operator will limit the minimum to zero. Accessibility can also be based on other approaches besides time depending on the context and should then be calculated by aggregating the different methods [Le06].

Believability

$$\begin{aligned} = min(&Believability\ of\ source;\\ &Believability\ when\ compared\ to\ internal\ commonsense\ standard;\\ &Believability\ based\ on\ the\ age\ of\ data\ ) \end{aligned} \tag{10}$$

The values of the three 'Believability' variables are rated from 0 to 1 by an individual who has background knowledge and can therefore evaluate the credibility of the source of the data. This metric for 'Believability' is based on subjective opinion and personal experience and is therefore open to biases and heavily context-dependent.

For DQD like 'Reputation', 'Interpretability' and others, specific metrics with a defined formula could not be found as they are very context-dependent and often subjective. The general approach for handling these DQD is that data quality rules (DQR) are defined, at best with the help of domain experts. The data sets are then checked for adherence to the specific rules and the number of data sets that violate the DQR are counted. That number is then divided by the total number of data sets verified by the rule and the result is subtracted from 1. A general example for the described approach is shown in equation (11).

$$1 - \frac{Number\ of\ data\ sets\ that\ violate\ DQR}{Number\ of\ data\ sets\ validated\ against\ DQR} \tag{11}$$

Since DQR constitute a separate part of DQ and a detailed discussion of the topic is beyond the scope of this paper, it is assumed in the following that the individual subjective dimensions can be captured via such a DQR in combination with equation 11. This assumption is based on literature that provides methods for the automatic discovery and assessment of such rules [Le06].

5 Data Quality Index

We will now synthesize the results of the previous chapters towards a DQI for data valuation in data economy. When analyzing the DQI, it was noticed that many DQD metrics are based on DQR. This is due to the fact that use cases are usually very specific and no generally applicable metrics exist, in contrast to dimensions like "Completeness", for example. Even with DQDs for which general metrics exist, it is usually assumed that it is defined beforehand what the individual dimensions mean within the application context. E.g., for the DQD 'Believability', it must be defined what makes a data set believable before the metric can be applied. For the DQI of this paper, we therefore conclude that in the first step, rules must be created for each of the DQD. For each DQD, at least one rule should exist. Guidelines and recommendations for the creation of such rules can already be found in the literature [Le18, Fa15].

Regarding DQD, it has been noticed that many existing DQIs consider very few DQD on average or use context-specific labels but refer to the aspects of the DQD from literature. This can be justified by the fact that the existing DQIs were mostly developed for a very narrowly defined use case. This approach is certainly sensible in order to make users comprehend what is assessed through each DQD. Since the goal of this paper is to design a DQI for data valuation in data economy, we use the general DQDs proposed in the literature to ensure the generalizability of the DQI [PLW02]. If more specific DQDs need to be created in individual cases, they must be assigned to one of the general DQDs afterwards to ensure manageability. As for the number of DQDs to be used, it will certainly be difficult to check every record for every DQD. Therefore, it seems advisable to make this decision depending on the data to be considered. In a hypothetical data marketplace scenario, this decision could be made by the buyer of the data as they define what constitutes high quality data in their use case. Since

the previously described study found only a slight tendency about the importance of each DQD, no DQD was excluded. However, a possible exclusion of DQDs would be justified by the fact that many of these DQDs are very subjective or context-dependent and require some domain knowledge to verify the correctness of the rules used.

Many of the existing DQIs provide little or inaccurate information regarding the metrics used. In some cases, these are also very context-specific and difficult to understand for non-experts. This is also reflected in the literature, as there are only a very limited number of publications that precisely develop metrics for specific DQDs or provide guidance on how these should be developed [He18, Fa15, Le06]. Our model uses the metrics of [Le06] in combination with general rule-based metrics for the context-dependent DQD. Weights in an index are a way to incorporate a certain dimension more strongly into the index or to make the index more dependent on one dimension. Within the scope of this work, an attempt was made to develop a possible weighting based on the ranking of the study. Since the differences between the individual DQDs were not sufficiently large to make a meaningful difference in the weighting, this approach was discarded and not pursued further. Reasons for this could be that the number of participants was not sufficient and a more extensive study is needed, or that the weights need to be considered in the context of different use cases. From the analysis of the existing approaches, a multitude of different methods emerges to determine a weighting. However, the basis of these weighting approaches are mostly estimations of domain experts or statistical calculations, which calculate a suitable weighting based on these estimations. For our model, we refrain from specifying a fixed weighting, as this should be determined by those who want to acquire the data. Thus, in the implementation it needs to be ensured that the party acquiring data is given an opportunity to customize the weighting of DQ. In addition, the sum of all weights must equal 1. Otherwise the aggregation equation will not provide the desired output of a value between 0 and 1. For the aggregation of individual values, possible weights are multiplied by the values of the DQD. The resulting weighted values are then summed up. The result is a value between 0 and 1, which represents the quality of the data considered. The higher the value, the higher the quality of the data. It is also possible to multiply this value by 100 to give the DQI as a percentage, if desired.

The resulting model is shown in Figure 1 which illustrates, how the implementation of a DQI can reduce the complexity of multi-dimensional models. As mentioned in the previous chapters, this reduction in dimension comes with an information trade-off. We were able to observe this trade-off in our analysis of already existing DQIs. They lacked comprehensibility in terms of how the DQI was calculated due to missing information either on the used rules, dimensions or metrics. Therefore we recommend to introduce a mandatory description of the data attributes, DQR, DQD and the metrics by the party that provides the data, when implementing this DQI model into valuation approaches for data economy.

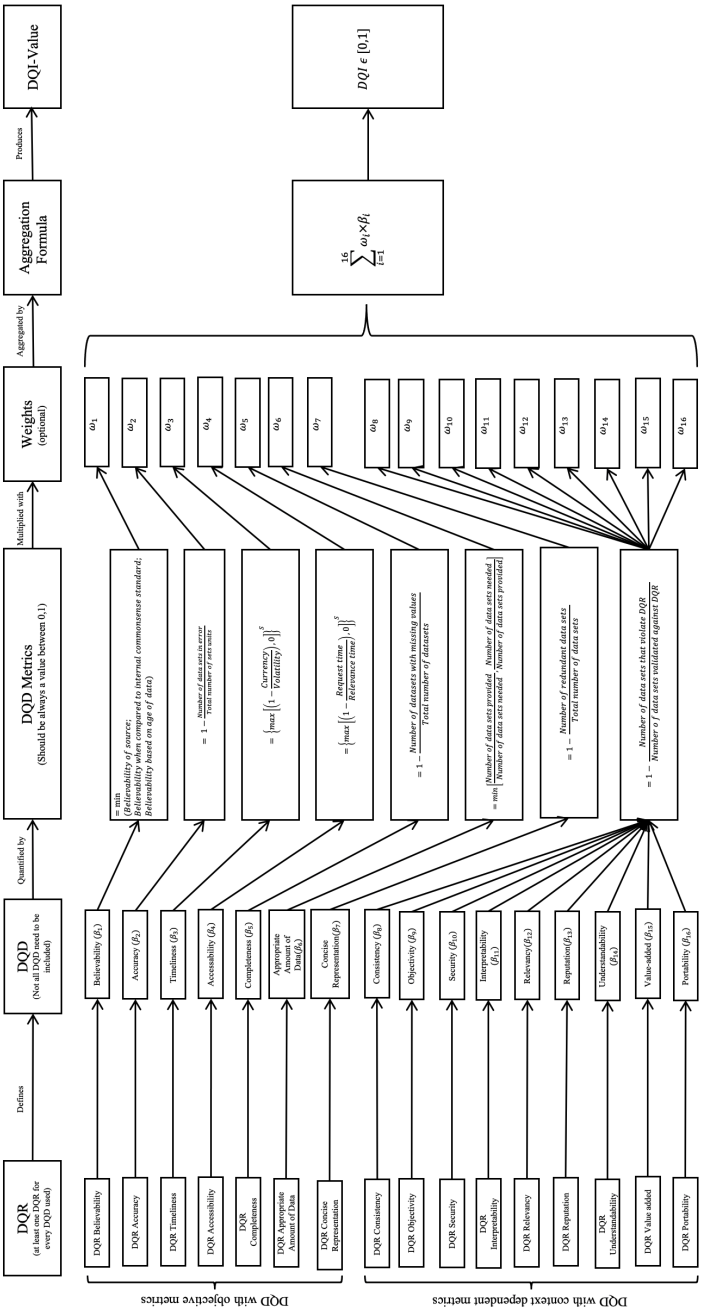


Fig. 1: General framework for developing a DQI

6 Conclusion

Our work does not come without limitations and room for improvement. First, we used the general approach to measure DQD. More precise formulas for the context-dependent DQD could be analyzed and implemented. While the general formula is mainly sufficient, easy to apply and provides a precise definition of DQRs, it cannot be guaranteed that this will do justice to every DQD in every use case. Similarly, this paper could not go further in-depth into the aspect of DQRs, which are a significant building block of the model. Although there are some approaches how these can be created, in the current model, however, the buyer of the data would have to check exactly whether the rules used also measure what they are supposed to. This requires domain knowledge and the availability of detailed descriptions, among other things. Since the study included only a small number of participants, the results are not necessarily representative and were not sufficient to develop a well-founded weighting of the DQD.

Future work should address these points and also examine approaches on how to automate the assessment process, how to incorporate the DQI to data valuation methods, and creating a service around the DQI. These points are especially important with regard to data economies, as the amount of data that will be shared or sold there will certainly not allow for manual valuation. A larger study could contribute to a reduction of the DQD and enable a more generalized weighting. The possible study could also consider different use cases in order to examine a possible difference in importance depending on the context.

In summary, in this paper we have gathered some basic arguments for a DQI and supported them with examples of existing indices from academia and practice. We then analyzed, compared, and evaluated these DQIs. Together with the results of a quantitative study, we have then incorporated these results into our own DQI model, which draws on metrics and suggestions from the literature. We therefore conceptualized a sturdy starting point for further research on DQIs for data valuation in the data economy.

7 Acknowledgement

This work is part-funded by the research project 'Future Data Assets' (grant number: 01MD19010C) funded by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) within the 'Smart Data Economy' technology program, managed by the DLR project management agency.

Bibliography

[Ab21] Abbas, Antragama Ewa; Agahari, Wirawan; van de Ven, Montijn; Zuiderwijk, Anneke; de Reuver, Mark: Business Data Sharing through Data Marketplaces: A Systematic Literature Review. J. Theor. Appl. Electron. Commer. Res., 16(7):3321–3339, 2021.

[ACS16] Anand, Abhijith; Coltman, Tim; Sharma, Rajeev: Four steps to realizing business value from digital data streams. MIS Q. Exec., 15(4), 2016.

[Ba98] Ballou, D.; Wang, R.; Pazer, H.; Tayi, G. K.: Modeling Information Manufacturing Systems to Determine Information Product Quality. Management Science, 44(4):462–484, 1998.

[Bo14] Borek, Alexander; Parlikad, Ajith Kumar; Woodall, Philip; Tomasella, Maurizio: A Risk Based Model For Quantifying The Impact Of Information Quality. Computers in Industry, 65(2):354–366, February 2014.

[CZ15] Cai, Li; Zhu, Yangyong: The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Sci. J., 14:2, 2015.

[DB16] Döring, Nicola; Bortz, Jürgen: Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften. Springer, Wiesbaden, 5 edition, 2016.

[EBH13] E.Baghi; B.Otto; H.Osterle: Controlling Customer Master Data Quality: Findings from a Case Study. 05 2013.

[EHO11] Ebner, Verena; Hüner, Kai; Otto, Boris: , Fallstudie Bayer CropScience AG: Entwurf und Implementierung geschäftsorientierter Datenqualitätskennzahlen, 01 2011.

[ES07] Even, Adir; Shankaranarayanan, G.: Utility-Driven Assessment of Data Quality. SIGMIS Database, 38(2):75–93, May 2007.

[Fa15] Fan, W.: Data Quality: From Theory to Practice. SIGMOD Rec., 44(3):7–18, dec 2015.

[Fa20] Faroukhi, Abou Zakaria; El Alaoui, Imane; Youssef, Gahi; Amine, Aouatif: Big data monetization throughout Big Data Value Chain: a comprehensive review. Journal of Big Data, 7:3, 01 2020.

[Ge18] Geuer, M.: , Prozessorientierter Data Quality Index erfolgreich einführen. https://www.business-information-excellence.de/datenqualitaet/91-prozessorientierter-data-quality-index-erfolgreich-einfuehren, 2018. [Accessed 2022-07-23].

[Ge21] Gelhaar, Joshua; Gürpinar, Tan; Henke, Michael; Otto, Boris: Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. Proc. of 25th PACIS, p. 121, 2021.

[Gl93] Glazer, Rashi: Measuring the Value of Information: The Information-Intensive Organization. IBM Systems Journal, 32:99 – 110, 02 1993.

[He18] Heinrich, B.; Hristova, D.; Klier, M.; Schiller, A.; Szubartowicz, M.: Requirements for Data Quality Metrics. J. Data and Information Quality, 9(2), 2018.

[Hi] Hickey, D.; O'Connor, R.; McCormack, P.; Kearney, P.; Roosa, R.; Brennan, R.: The Data Quality Index: Improving Data Quality in Irish Healthcare Records. [Hi], pp. 625–636.

[HK15] Hönigl, Jürgen; Küng, Josef: Obtaining a data quality index with respect to case bases. Vietnam Journal of Computer Science, 2(1):47–56, 2015.

[HN20] Hanafizadeh, Payam; Nik, Mohammad Reza Harati: Configuration of Data Monetization: A Review of Literature with Thematic Analysis. Global Journal of Flexible Systems Management, 21(1):17–34, 2020.

[In21] International Aid Transparency Iniative: , IATI Data Quality Index. https://prod-iati-website.azureedge.net/prod-iati-website/documents/IATI_Data_Quality_Index_-_Background_Paper_1_.pdf, 2021. [Accessed 2022-07-23].

[Ir21] Ireland, Sheep: , Sheep Ireland Guide and Directory of Breeders. https://www.sheep.ie/wp-content/uploads/2021/08/SHEEP-IRELAND-GUIDE-DIRECTORY-2021-reading-view.pdf, 2021. [Accessed 2022-07-23].

[Jo18] Journal, World Economics: , The Data Quality Index (DQI). https://www.world-economics-journal.com/Pages/Data-Quality-Index.aspx, 2018. [Accessed 2022-07-23].

[JT18] JT Lorenz: . https://www.mckinsey.com/industries/financial-services/our-insights/the-rise-of-ecosystems-and-platforms-what-role-can-insurers-play-and-how-can-they-get-started, 2018. [Accessed 2022-07-23].

[Ka17] Karam, S.: , International trade data quality index. http://researcharchive.vuw.ac.nz/handle/10063/6128, 2017. [Accessed 2022-07-23].

[KSW02] Kahn, Beverly K.; Strong, Diane M.; Wang, Richard Y.: Information Quality Benchmarks: Product and Service Performance. Commun. ACM, 45(4):184–192, April 2002.

[Le02] Lee, Yang W.; Strong, Diane M.; Kahn, Beverly K.; Wang, Richard Y.: AIMQ: a methodology for information quality assessment. Information Management, 40(2):133–146, 2002.

[Le06] Lee, Y. W.; Wang, R. Y.; Funk, J.D.; Pipino, L. L.: Journey to Data Quality. MIT Press, Cambridge, 2006.

[Le18] Lee, E.; Park, Y.; Kim, H.; Jung, B.; Hong, J.: An Efficient Method for Automatic Updating Business Rules. In: Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems. RACS '18, Association for Computing Machinery, New York, NY, USA, p. 331–332, 2018.

[MW99] Moody, Daniel L.; Walsh, Peter: Measuring the Value Of Information - An Asset Valuation Approach. In: ECIS. 1999.

[Na20] Naik, Shanoja; Voong, Stephanie; Bamford, Megan; Smith, Kyle; Joyce, Angela; Grinspun, Doris: Assessment of the Nursing Quality Indicators for Reporting and Evaluation (NQuIRE) database using a data quality index. Journal of the American Medical Informatics Association : JAMIA, 27:776–782, 05 2020.

[NS17] Nagle, Tadhg; Sammon, David: The Data Value Map: a Framework for Developing Shared Understanding on Data Initiatives. In: ECIS. 2017.

[ORH05] Oliveira, Paulo; Rodrigues, Fátima; Henriques, Pedro Rangel: A Formal Definition of Data Quality Problems. In: ICIQ. 2005.

[OW08] Otto, Boris; Weber, Kristin: Data Governance. volume 53. pp. 265–283, 01 2008.

[PA22] Data Quality Index. https://www.paconsulting.com/industries/energy-and-utilities/data-quality-index/, [Accessed 2022-07-23].

[PLW02] Pipino, Leo L.; Lee, Yang W.; Wang, Richard Y.: Data Quality Assessment. Commun. ACM, 45(4):211–218, April 2002.

[Re19] Regneri, Michaela; Georgi, Juliane; Kost, Jurij; Pietsch, Niklas; Stamm, Sabine: Computing the Value of Data: Towards Applied Data Minimalism. ArXiv, abs/1907.12404, 2019.

[SB16] https://www.bankingsupervision.europa.eu/press/conferences/shared/pdf/sup_rep_conf/2017/Data_quality_framework_tools_and_products.pdf, [Accessed 2022-07-23].

[Se16] Serhani, Mohamed Adel; Kassabi, Hadeel T. El; Taleb, Ikbal; Nujum, Al Ramzana: An Hybrid Approach to Quality Evaluation across Big Data Value Chain. 2016 IEEE International Congress on Big Data (BigData Congress), pp. 418–425, 2016.

[SHS13] Selander, L.; Henfridsson, O.; Svahn, F.: Capability Search and Redeem across Digital Ecosystems. Journal of Information Technology, 28(3):183–197, 2013.

[SM22] Stein, Hannah; Maaß, Wolfgang: Requirements for Data Valuation Methods. In: 55th Hawaii International Conference on System Sciences. Hawaii International Conference on System Sciences (HICSS-2022), January 4-7, Hawaii, Hawaii, United States. Springer, 2022.

[ST17] Short, James E.; Todd, Steve: , What's Your Data Worth? — sloanreview.mit.edu. https://sloanreview.mit.edu/article/whats-your-data-worth/, 2017. [Accessed 09-Nov-2021].

[SWL20] Shi, Q.; Wang, H.; Lu, H.: Research and Application of AHP-EWM-Based Comprehensive Evaluation of Data Quality. In: Proceedings of the 4th International Conference on Computer Science and Application Engineering. CSAE 2020, Association for Computing Machinery, New York, NY, USA, 2020.

[Ta16] Taleb, Ikbal; Kassabi, Hadeel T. El; Serhani, Mohamed Adel; Dssouli, Rachida; Bouhaddioui, Chafik: Big Data Quality: A Quality Dimensions Evaluation. 2016 Intl IEEE Conferences on Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), pp. 759–765, 2016.

[Un22] United Nations Development Fund: . https://hdr.undp.org/data-center/human-development-index#/indicies/HDI, 2022. [Accessed 2022-07-23].

[Wi12] Wirtz, Harald: Valuation of Intellectual Property: A Review of Approaches and Methods. International Journal of Biometrics, 7:40, 2012.

[WML01] Wang, Richard Y.; Mostapha, Ziad; Lee, Yang W.: Data quality. Kluwer Academic Publishers, Boston, 2001.

[WS96] Wang, Richard Y.; Strong, Diane M.: Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems, 12(4):5–33, 1996.

[WS22] Woroch, Robert; Strobel, Gero: Show me the Money: How to monetize data in data-driven business models? 02 2022.

Appendix

Origin	Data Type	Use Case	DQD	Strength	Weakness
[Ir21]	Agriculture	DQI for ranking sheep and evaluating their monetary value based on breeding data and its accuracy.	Accuracy, Timeliness, Completeness	Gives an example for linking data quality to monetary values.	DQD are not separated clearly
[HK15]	Type-independent	DQI for assessing cases in a case-based reasoning (CBR) system	Average solutions per case[3]; Count of similar retained queries[4]; Missing Values[5]	Importance & usefulness of DQI for AI paradigms; Specific DQ criteria for CBR mapped to general DQD.	Same DQD are measured by different criteria
[Ka17]	Trading	DQI to assess consistency in regards to bilateral trade data reporting between countries.	Data availability, Import/ export size ratio and share	Integrate different sources for the same information to detect inconsistencies in data sets.	Focus on "consistency" dimension and provides less insights into other DQD.
[Hi]	Healthcare	DQI for Irish healthcare system records	Correctness of entity specific attributes	Proof that DQI implementation leads to a system wide increase in data quality	Rules merely assess the record for missing values; Unclear if and how other DQD are assessed.

Tab. 4: Overview of the existing DQIs from literature and practice - part 2

[3] reflects the following DQD:appropriate amount of data, concise representation, ease of manipulation, free-of-error, interpretability, timeliness, understandability, value added

[4] reflects the following DQD:appropriate amount of data, free-of-error, interpretability, timeliness, value added

[5] reflects the following DQD:appropriate amount of data, concise representation, completeness, free-of-error, interpretability, relevancy, understandability, value added

Origin	Data Type	Use Case	DQD	Strength	Weakness
[Ir21]	Agriculture	DQI for ranking sheep and evaluating their monetary value based on breeding data.	Accuracy, Timeliness, Completeness	Gives an example for linking data quality to monetary values.	DQD are not separated clearly
[PA22]	Utility Industry	DQI for monitoring and improving DQ. Uses business rules on data assets linked to certain DQD which are weighted and aggregated into a DQI.	Accuracy, Completeness, Consistency, Uniqueness, Validity, Integrity, Timeliness	Provide clear structure for DQI using business rules and connecting them to DQD.	Lacks basics for metrics and their application.
[Ge18]	Process	Monitoring company processes against a desired DQI target.	Timeli-, Unique-, Correct-, Completeness, Accuracy, Consistency, Non-redundancy, Relevancy, Uniformity, Reliability, Understandability	A high number of DQD are included into the DQI; Metrics are explained comprehensively.	Missing literature on how the metrics are being derived.
[Jo18]	Finance	DQI to measure the quality of income data of 154 countries to rank them by their trustworthiness.	Base year, System of national accounts, The informal Economy, Quality of statistics, Corruption	Example of how DQI can be used to establish trust when data is self-reported.	DQI is based on other indices which increases complexity and decreases transparency.
[In21]	Government	DQ Framework based on a DQI to create a DQ standard for transparent data sharing.	Timeliness, Comprehensiveness, Forward-Looking	Continuous feedback rounds for DQI improvement.	Calculation of the DQD values lacks comprehensibility due to domain-specific key figures.

Tab. 5: Overview of the existing DQIs from literature and practice - part 3

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1055

Teilautomatisierte Datenqualitätsbewertung und Fehlerkorrektur zur Senkung der Einstiegshürde von Datenanalysemethoden

Philipp Schlunder[1]

Abstract: Zur Nutzung aktueller Methoden der Datenanalyse müssen Daten häufig zunächst aufbereitet werden. Hierzu werden Kenntnisse benötigt, die in viele Organisationen noch nicht vorliegen. Um diese Einstiegshürde zu verringern, wird ein Konzept zur interaktiven, teilautomatisierten Behebung von Datenqualitätsproblemen vorgestellt, dessen Anwendung weniger Vorkenntnisse bedarf, indem eine automatische Bewertung der Datenqualität und des Informationsgehalts bereitgestellt wird, mit der Option, erkannte Probleme durch verschiedene Ansätze direkt korrigieren zu lassen.

Keywords: Datenqualität; Maschinelles Lernen; KMU; Einstiegshürde

1 Motivation und Stand der Technik

Deutsche KMU[2] stellen knapp 99 % der Unternehmen dar, doch nur 6 % von ihnen nutzen bereits Künstliche Intelligenz (KI) [BMWi]. Knapp über 50 % der Unternehmen nennen fehlende Expertise und knapp 28 % fehlende finanzielle Ressourcen als Grund für den mangelnden Einsatz [BVMW20]. In den meisten Fällen müssen Daten vor einer Analyse aufbereitet werden. Dazu gehört die Bewertung der vorliegenden Datenqualität, sowie des Informationsgehaltes und damit die Eignung für eine anschließende Analyse. Bestehende Methoden zur Qualitätsbewertung [Af20; Ei19; Gu21], die teils interaktiv und detailliert Probleme aufzeigen (Pandas-Profiling [PP], Tensorflow Data Validator [Ma15]), setzen Programmierkenntnisse voraus, sind eingeschränkt auf eine vordefinierte Korrekturmaßnahme pro Fehler (Great Expectations [Go]) oder erlauben Fehlerkorrekturen über ein GUI vorzunehmen, bieten aber keine Handlungsempfehlung an (RapidMiner TurboPrep [RMTP18]). Im Vergleich der bewerteten Werkzeuge wird deutlich, dass aktuell zur Bewertung und Behebung von Datenqualitätsproblemen, sowie zur Einschätzung des Informationsgehaltes eines Datensatzes entweder Programmierkenntnisse oder Kenntnisse über Datenanforderungen zum Einsatz von KI benötigt werden.

[1] daibe UG, pschlunder@daibe.io

[2] Klein und mittelständische Unternehmen mit weniger als 250 Mitarbeitenden

doi:10.18420/inf2022_87

2 Konzept

Zielgruppe des vorgestellten Konzepts sind Personen einer Organisation, die über Domänenwissen verfügen, aber über geringe bis keine Programmier- und Datenanalysekenntnisse. Sie besitzen lediglich ein erstes Grundverständnis vom Zusammenhang zwischen Daten und abgebildeten Informationen (Data Literacy). Das Konzept soll diese Personen befähigen, eigenständig ein erstes Verständnis über die Qualität und Eignung von Daten für weitere Analysezwecke zu erlangen, damit sie eigenständig Fehler im betrachteten Datensatz informiert beheben können. Um dieses Bewusstsein herzustellen, werden Informationen zur erhobenen Datenqualität in verschiedenen Granularitätsstufen bereitgestellt, und unterschiedliche Behebungsmechanismen angeboten, die zusammen mit einer Erklärung, Nutzenden erlauben Korrekturen an geladenen Datensätzen vorzunehmen.

Das Konzept sieht eine Web-App vor, die Personen vom Einlesen ihrer Daten über eine Bewertung und optionalen Fehlerkorrektur der Datenqualität bis hin zu einer Anschlussverwertung begleitet. Im Kern besteht der Prozessfluss des Konzepts aus folgenden Schritten:

1. Exemplarische Daten per Drag & Drop anbinden.
2. Darstellung eines Datenauszugs mit automatischer Erkennung vorliegender Datentypen und Identifikation einer als ID geeigneten Spalte.
3. Auswahl der zu analysierenden Teilmenge der Daten.
4. Bereitstellung einer Bewertung der Datenqualität, sowie des Informationsgehalts in allgemeiner und detaillierter Form.
5. Möglichkeit zur (teil-)automatischen Korrektur von Datenqualitätsproblemen.
6. Bereitstellung reparierter und entfernter Daten, sowie Qualitäts- und Änderungsberichte. Option, durchgeführte Korrekturlogik auf neuen Daten anzuwenden, indem je nach Wunsch eine Web-App, ein REST-API oder eine CLI bereitgestellt wird.

Die Bereitstellung als Web-App mit der Möglichkeit, Daten in gängigen Formaten via Drag & Drop zu laden, ermöglicht es auch nicht IT-affinen Personen, das Werkzeug einfach bedienen zu können. Bei jedem der Schritte ist Nutzenden freigestellt, ob sie Automatisierungen verwenden oder manuell eingreifen, indem Sie beispielsweise selbst eine ID-Spalte auswählen; entscheiden, ob und wie ein Datenqualitätsproblem behoben wird; oder der Überprüfungskontext angepasst wird, um regionsspezifische Einstellungen und durchzuführende Qualitätsüberprüfungen zu verändern.

Insbesondere das Laden heterogener, teilstrukturierter Dateien, sowie die Auswahl von Fehlerkorrekturen wird durch verschiedene Automatisierungen und Einstellungsmöglichkeiten unterstützt. Um zu gewährleisten, dass reale Daten eingebunden werden können, wird beispielsweise das Encoding einer Datei erkannt und bei XLSX-Dateien eine Bereichserkennung auf Datenblättern durchgeführt. Bei CSV-Dateien wird der eigentliche Zeilenbereich

herausgefiltert, welcher oft durch Meta-Informationen zu Beginn oder Ende der Dateien umgeben ist.

Datenqualitätsprobleme werden zusammenfassend beschrieben aufgeführt, mit der Option, detaillierte statistische Beschreibungen, sowie Visualisierungen der Problematik anzeigen zu lassen; einen Fehler zu ignorieren; eine automatische Korrektur zu verwenden, oder diese manuell anzupassen. Dazu werden verschiedene Korrekturmechanismen angeboten, wobei der empfohlene ausgewählt ist und gängige Einstellungen anpassbar sind. Jedoch sind die Parameterbeschreibungen in eine anwendergerechte Sprache überführt und mit dem Datensatz entsprechenden Parametern vorbefüllt. Je nach Fehlertyp wird eine Vorauswahl getroffen, ob ein Fehler behoben werden soll oder nicht. Somit ist es Nutzenden möglich, nach der Überprüfung eine Korrektur automatisch mit vorgeschlagenen Mechanismen vorzunehmen, oder diese anzupassen. Bereitgestellte Informationen umfassen dabei immer auch Auflistungen problematischer Einträge, während Korrekturempfehlungen umgangssprachliche Beschreibungen des Mechanismus bereitstellen.

Durch die Befähigung zur Qualitäts- und Informationsgehaltsbewertung können Nutzende so bestehende Datensätze anpassen und vergleichen, um eine Vorauswahl an Daten zu erhalten, die für ein etwaiges Analyseprojekt verwendet werden können. Diese Vorarbeiten können dabei helfen, anfängliche Aufwände zur Datensammlung und -Vorverarbeitung durch Expert*innen zu reduzieren, und somit Kosten zusenken. Zudem lernen Nutzende, welche potenziellen Fehler in Daten auftreten können und was für Korrekturmechanismen existieren, um so gegebenenfalls interne Datenerhebungsprozesse proaktiv anzupassen.

3 Technische Umsetzung

Es existiert bereits eine erste Umsetzung des Konzeptes in Python, wobei `streamlit`[3] zur Erstellung des Front-Ends eingesetzt wird, und das Back-End auf einer neu entwickelten Bibliothek zur Überprüfung und Korrektur von Datenqualitätsproblemen aufbaut. Im Folgenden wird der Aufbau der Bibliothek beschrieben. Zentrale Konzepte sind:

- **CheckContext**: Verwaltet Meta-Informationen zur Konfiguration einer Qualitätsprüfung. Dies beinhaltet eine Liste an durchzuführenden *QualityChecks*, samt Voreinstellungen, sowie regionale Einstellungen, wie die typische Darstellung von Zeitstempeln, der Zeitzone, Währungseinstellungen und der Zielsprache.
- **DataQualityHandler**: Lädt einen *CheckContext*, wendet die definierten *QualityChecks* auf bereitgestellte Daten an und verwaltet erkannte Probleme und ausgewählte Korrekturen. Vor der Anwendung der Checks auf Daten werden diese bedingt durch regionale Einstellungen vorkonfiguriert.
- **QualityChecks**: Mechanismen zur Überprüfung einzelner Datenqualitätsmetriken, die vordefinierte *QualityFixes* enthalten.

[3] Python-Bibliothek zur Erstellung von Web-Apps aus Analyse-Skripten, `https://streamlit.io/`

- **Warnings**: Bei Anwendung eines Checks auf Daten werden Qualitätsmetriken erhoben, *QualityFixes* angepasst und ein empfohlener Fix vorausgewählt. Dabei werden regionale Einstellungen aus der Vorkonfiguration durch den *DataQualityHandler* berücksichtigt. Diese datenbezogenen Ergebnisse eines *QualityChecks* werden als *Warning* abgespeichert.
- **QualityFixes**: Sind Korrekturmaßnahmen zur Behebung von Datenqualitätsproblemen, wobei ein Fix für mehrere Checks als Maßnahme benutzt werden kann. *QualityFixes* enthalten allgemeinverständliche Beschreibungen ihrer Methodik und Parametrisierung.
- **Reports**: *DataQualityHandler* sammeln generierte *Warnings* in einem *Report*, welcher alle durchgeführten *QualityChecks* und angewandten *QualityFixes* verwaltet. Diese können als textueller Bericht und Konfigurationsdatei ausgegeben werden.

Es existieren bereits Datenqualitätschecks und Korrekturmechanismen zur Handhabung von Problemen der Vollständigkeit, Konsistenz kategorischer Ausprägungen, unerwarteter Sprungstellen und doppelter Einträge. Jeder Datenqualitätscheck berechnet verschiedene Kennzahlen zur Bewertung, ob ein Problem vorliegt, und welche Methode zur Korrektur vorgeschlagen wird. Eine Kombination dieser Metriken wird zudem verwendet, um eine einzelne, prozentuale Kennzahl als erstes Maß für die Gesamtqualität anzubieten. Der Informationsgehalt wird aktuell über die Entropie, die Korrelation der Spalten untereinander, sowie der Güte der (erkannten) ID-Spalte bestimmt. Dabei wird angenommen, dass eine höhere Entropie und eine geringe Korrelation zwischen Spalten im Bezug auf ihre Anzahl ein höheres Potenzial zum Finden von Strukturen in den Daten bietet, sowie berücksichtigt, dass in den meisten Analyse-Szenarien eine konsistente ID benötigt wird.

4 Zusammenfassung und Ausblick

Zur Reduktion der Einstiegshürde in die Nutzung von Datenanalyse und KI wurde ein Konzept zur Überprüfung und Korrektur der Qualität von Datensätzen vorgestellt, welches es Personen ohne Programmier- und Datenanalyse-Kenntnissen ermöglichen soll, vorhandene Daten zu beurteilen und Qualitätsprobleme zu beheben. Dazu wurde das Konzept als prototypische Web-App basierend auf einer neuen Bibliothek zur Datenqualitätshandhabung umgesetzt, um teilautomatisierte Korrekturmechanismen bereitstellen zu können. Nutzende erhalten so die Möglichkeit, eigenständig und informiert Fehler zu erkennen, zu beheben und korrigierte Daten, sowie neu gewonnene Erkenntnisse als Start für Analysen oder Prozessanpassungen in ihrer Organisation zu verwenden.

In Zukunft wird die Bibliothek als open-source Lösung bereitgestellt und die Integration bestehender Datenvalidierungsframeworks geprüft.

Literatur

[Af20] Afzal, S.; C, R.; Kesarwani, M.; Mehta, S.; Patel, H.: Data Readiness Report./, Okt. 2020, URL: http://arxiv.org/abs/2010.07213.

[BMWi] Bundesministerium für Wirtschaft und Energie: Einsatz von Künstlicher Intelligenz in der Deutschen Wirtschaft, URL: https://www.bmwk.de/Redaktion/DE/Publikationen/Wirtschaft/einsatz-von-ki-deutsche-wirtschaft.pdf?__blob=publicationFile&v=8, Stand: 25. 07. 2022.

[BVMW20] Der Mittelstand, BVMW e. V.: Digitalisierung im Mittelstand: Aufschwung, aber kein Durchbruch, 2020, URL: https://www.bvmw.de/themen/digitalisierung/news/12285/digitalisierung-im-mittelstand-aufschwung-aber-kein-durchbruch/, Stand: 25. 07. 2022.

[Ei19] Eickelmann, M.; Wiegand, M.; Deuse, J.; Bernerstätter, R.: Bewertungsmodell zur Analyse der Datenreife. Zeitschrift für wirtschaftlichen Fabrikbetrieb 114/, S. 29–33, Feb. 2019, ISSN: 2511-0896, URL: https://www.degruyter.com/document/doi/10.3139/104.112037/html.

[Go] Gong, A.; Campbell, J.; Superconductive; Great Expectations: Great Expectations, URL: https://github.com/great-expectations/great_expectations.

[Gu21] Gupta, N.; Patel, H.; Afzal, S.; Panwar, N.; Mittal, R. S.; Guttula, S.; Jain, A.; Nagalapatti, L.; Mehta, S.; Hans, S.; Lohia, P.; Aggarwal, A.; Saha, D.: Data Quality Toolkit: Automatic assessment of data quality and remediation for machine learning datasets./, Aug. 2021, URL: http://arxiv.org/abs/2108.05935.

[Ma15] Martín Abadi; Ashish Agarwal; Paul Barham; Eugene Brevdo; Zhifeng Chen; Craig Citro; Greg S. Corrado; Andy Davis; Jeffrey Dean; Matthieu Devin; Sanjay Ghemawat; Ian Goodfellow; Andrew Harp; Geoffrey Irving; Michael Isard; Jia, Y.; Rafal Jozefowicz; Lukasz Kaiser; Manjunath Kudlur; Josh Levenberg; Dandelion Mané; Rajat Monga; Sherry Moore; Derek Murray; Chris Olah; Mike Schuster; Jonathon Shlens; Benoit Steiner; Ilya Sutskever; Kunal Talwar; Paul Tucker; Vincent Vanhoucke; Vijay Vasudevan; Fernanda Viégas; Oriol Vinyals; Pete Warden; Martin Wattenberg; Martin Wicke; Yuan Yu; Xiaoqiang Zheng: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, Software available from tensorflow.org, 2015, URL: https://www.tensorflow.org/.

[PP] Pandas Profiling, URL: https://github.com/ydataai/pandas-profiling, Stand: 25. 07. 2022.

[RMTP18] RapidMiner TurboPrep, 2018, URL: https://docs.rapidminer.com/9.3/studio/turbo-prep/, Stand: 24. 07. 2022.

Prescriptive and descriptive quality metrics for the quality assessment of operational data

Quality assessment for data-driven and hybrid models in the process industry

Isabell Viedt[1], Jonathan Mädler[1], Valentin Khaydarov[2] and Leon Urbas[1]

Abstract: In the process industry data-driven and hybrid modeling approaches are increasingly popular in regards to process monitoring, optimization and control. The major problem with process data is that the data collected in process plants during operation, even though available in vast amounts, might generally be low in information content. The collected data usually represents certain operating points while anomalies, ramp-up and shut-down are rare occurrences and therefore only seldom covered. Due to its possibly low quality, the use of such data might lead to an inadequate model coverage and overall low model performance. Data quality assessment prior to modeling is crucial to allow an estimation of model quality prior to the model development. Therefore, the following paper discusses prescriptive and descriptive assessment metrics for the quality assessment of process data and their potential application in the quality assurance of data-driven and hybrid models. This approach will in later application support the user in their choice of modeling approach.

Keywords: Data quality assessment, prescriptive data quality metrics, hybrid modeling.

1 Introduction

In the process industry most process cannot sufficiently be described via a first principles approach because most times there is a lack of knowledge about the underlying biochemical, chemical and biological process [CR19]. Therefore, data-driven and hybrid modeling approaches are increasingly popular in the process industry in regards to process monitoring, optimization and control [St14, Gär21]. Data collected during plant operation usually only covers certain operation point and can therefore be low in information content. Special events as anomalies, ramp-up and shut down are only seldom covered [Me19]. The use of low quality process data might lead to an inadequate model coverage and overall low model performance. This is especially important the more data is eventually used for the implementation of these simulation models. Hybrid models will generally require less over all data and lower quality data for the same performance as solely data-driven models [St14]. Even though the required data quality varies for the individual modeling application, for both hybrid and data-driven models the resulting

[1] Technische Universität Dresden, Professur für Prozessleittechnik, Helmholtzstr. 10, 01069 Dresden, isabell.viedt@tu-dresden.de, jonathan.maedler@tu-dresden.de, leon.urbas@tu-dresden.de
[2] Technische Universität Dresden, Process-to-Order-Lab (P2O-Lab), Helmholtzstraße 16, 01069 Dresden, valentin.khaydarov@tu-dresden.de

doi:10.18420/inf2022_88

model performance correlates with the utilized data for modeling. As data quality of process data is generally lower, it is important to consider data quality assessment prior to model development. With this, unsuitable or insufficient data can either be flagged for pre-processing to improve the overall quality or discarded all together because the data utilization would lead to insufficient model performance [St14]. To address this challenge, the following paper discusses how prescriptive assessment metrics can be utilized in the assessment of process data to support the user in decision making and their potential application in the quality assurance of data-driven and especially hybrid models [VMU22].

2 Data quality assessment

In order to enable informed decision making about the quality of data, it is first necessary to define what good data quality dimensions are [MVU21]. Therefore, the authors utilize already existing data quality frameworks that define attributes that can be allocated to data quality. One example for a data quality model to describe data quality dimensions is the quality model of ISO/IEC 25012 [ISO08]. This quality model is commonly used for the data quality assessment in software. The standard differentiates between two types of data quality: inherent data quality and system-dependent data quality.

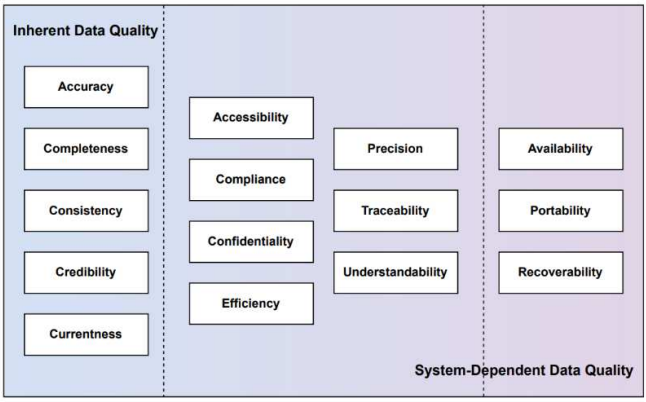


Fig. 1: Data quality model according to ISO/IEC 25012

Inherent data quality refers to the degree to which quality characteristics of data have the intrinsic potential to satisfy stated and implied needs. Therefore, data quality refers to data itself in this case. Whereas system-dependent data quality refers to the degree to which data quality is reached and preserved within a computer system when data is used under specified conditions and therefore quality depends on the used technological domain.

Besides dimensions for data quality, in literature already exist numerous frameworks for the assessment of data quality [CR19, Ja20]. Most only consider the data quality

dimensions or characteristics and disregard actually assessment measures and metrics in favor of being generic and widely applicable [CR19, He18]. To allow qualitative and quantitative assessment of the data quality dimensions, specific measures or metrics must be allocated to them in form of statistical measures and assessment specifications [He18, CR19, Ja20]. The SQuaRE standard already provides an initial assessment approach with the accompanying standard ISO/IEC 25024 [ISO15] which defines data quality measures (DQM) for quantitatively measuring the data quality in terms of characteristics defined in ISO/IEC 25012. In the following, the term data quality metrics is used to define the DQM.

3 Descriptive and prescriptive data quality metrics

With the quality process data being poor in most cases, strategies for pre-assessing the quality of the data before model implementation are necessary [MVU21]. For this purpose, the authors propose the structuring of the data quality assessment measures or metrics into the categories prescriptive data quality metrics and descriptive data quality metrics similar to prescriptive and descriptive data analytics [Me19]. Prescriptive data quality metrics allow the assessment of the data quality before implementation into a model or algorithm, descriptive metrics will assess the quality of the implemented software or modeling product. Descriptive data quality assessment (cf. descriptive analytics [Me19]) will give insights into the overall data quality. With this definition most quality metrics can at some point be defined as descriptive metrics. When analyzing ISO/IEC 25024 in regards to the defined data quality measures, it can be concluded that most system-dependent and inherent quality metrics can be classified as descriptive metrics. The problem with a descriptive approach is that there is no recommended action and the fault detection only occurs after model implementation. Therefore, model improvement through an additional iteration loop is required, which utilizes more resources, time, and money.

The concept of prescriptive data quality assessment has the purpose of assessing data quality prior to model implementation and also includes recommendations of action in regard to type of suitable pre-processing steps and modeling approach. With this, the modeling options are narrowed down to the most suitable methods. To achieve prescriptive data quality assessment, specific data quality metrics and measures must be allocated. One example of a prescriptive data quality measure is the measure ‘data value completeness’ allocated to the inherent quality dimension completeness. The purpose of the measure and metric is to verify if the data set for example represents all necessary features or operating points. Further ‘Data model accuracy’ for the quality dimension accuracy coupled with a statistical approach can be utilized to predict model quality via data quality [He18]. Prescriptive data quality assessment is an interesting addition to the quality- and test-driven modeling approach presented in Mädler et al. (2021) [MVU21]. This approach also requires the quality and scope of the necessary context data to assess the inherent data quality prescriptively [VMU22]. Therefore, qualitative measures like ‘Metadata completeness’ for the quality dimension completeness [ISO15] can also be utilized in prescriptive data quality assessment.

4 Conclusion and outlook

This paper presented the concept of structuring data quality metrics into a prescriptive and descriptive category. With this, the first step into integrating this data quality assessment approach into a general model quality assessment framework (cf. [VMU22, MVU21]) has been taken. Since data quality is important for the quality assessment of hybrid models, future assessment framework can be improved by conducting prescriptive data quality assessment before the start of model building to exclude unfit data or to select the most suitable data set. This will also allow a supported decision between different data-driven and hybrid modeling approaches based on the overall data quality. The next step is the validation via case studies. For this quality models for the specific modeling goal are derived and the data is then tested against those quality models, metrics and reference values to verify if the presented approach supports the user with model selection.

Literaturverzeichnis

[CR19] Cichy, C.; Rass, S.: An overview of data quality frameworks. IEEE Access 7, 24634-24648, 2019.

[Gär21] Gärtler, M. et al.: The Machine Learning Life Cycle in Chemical Operations–Status and Open Challenges. Chemie Ingenieur Technik, 93(12), 2063-2080, 2021.

[He18] Heinrich, B. et.al.: Requirements for data quality metrics. Journal of Data and Information Quality (JDIQ) 9(2), 1-32, 2018.

[ISO08] ISO/IEC 25012:2008, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Data quality model, 2008.

[ISO15] ISO/IEC 25024:2015, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Measurement of Data Quality, 2015.

[Ja20] Jain, A. et al.: Overview and importance of data quality for machine learning tasks. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 3561-3562, 2020, August.

[Me19] Menezes, B. C. et al.: Predictive, prescriptive and detective analytics for smart manufacturing in the information age. IFAC-PapersOnLine, 52(1), 568-573, 2019.

[MVU21] Mädler, J., Viedt, I., Urbas, L.: Applying quality assurance concepts from software development to simulation model assessment in smart equipment. In Computer Aided Chemical Engineering. Elsevier. 50, 813-818, 2021.

[St14] Von Stosch, M., Oliveira, R., Peres, J., de Azevedo, S. F.: Hybrid semi-parametric modeling in process systems engineering: Past, present and future. Computers & Chemical Engineering, 60, 86-101, 2014.

[VMU22] Viedt, I., Mädler, J., Urbas, L.: Quality assessment for dynamic, hybrid semi-parametric state observers. In Computer Aided Chemical Engineering. Elsevier. 51, 813-818, 2022.

Agile Methoden im Informatikunterricht

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1067

Workshop zum Einsatz agiler Methoden im Informatikunterricht

Ein Minigolf-Spiel als erstes Softwareprojekt

Tobias Bahr[1] und Bernd Zinn[2]

Abstract: In diesem Bericht zum Workshop für Informatiklehrpersonen wird eine Unterrichtseinheit zum Themenbereich Algorithmen der Klassenstufe 8 für ein Programmierprojekt mit Scratch vorgestellt. Die teilnehmenden Informatiklehrpersonen sollen in dem Workshop die Grundlagen des agilen Prozesses lernen und reflektieren unterrichtsrelevante Aspekte, indem diese die Schülerperspektive einnehmen sowie strukturgebend die Vor- und Nachteile zum Einsatz von agilen Methoden im Informatikunterricht diskutieren. Abschließend werden Erkenntnisse aus der Unterrichtspraxis vorgestellt sowie methodisch-didaktische Aspekte zur Umsetzung von Agilen Methoden diskutiert.

Keywords: Informatikunterricht; Agile Methoden; Projekt; Softwareentwicklung; kollaboratives Arbeiten

1 Motivation

Agile Methoden werden bereits seit mehreren Dekaden im wirtschaftlichen Sektor und insbesondere auch in der Softwareentwicklung eingesetzt [Co05, Wi03]. Über die Jahre haben sie auch ihren Weg in den Informatikunterricht der Schulen gefunden [Br15, Br19, Di15 Ka15, Ri09]. Anknüpfend an die pädagogische und fachdidaktische Forschung zu der Projektmethode [Sc13] wurden praxisnahe Handlungsanweisungen für die unterrichtliche Umsetzung formuliert [Wa09, Sc13, Tr11]. Wenngleich sich die vorliegenden Veröffentlichungen zu agilem Informatikunterricht bislang nur mit der unterrichtlichen Umsetzung und kontextuellen Einkleidung beschäftigen, nach den uns vorliegenden Daten keine Evidenz zu den Wirkungseffekten der Umsetzung der Methode vorliegt, wird übergreifend unterstellt, dass die Nutzung agiler Methoden im Unterricht das Potenzial hat, um die personalen und sozialen Kompetenzen (u. a. Kommunikationsfähigkeit, Selbstorganisationfähigkeit, Teamfähigkeit) von Schüler*innen zu fördern [Br19, Ri09]. Im hier thematisierten Workshop wird eine Unterrichtseinheit vorgestellt, mit der im Anfangsunterricht Informatik anhand eines Softwareprojekts für die 8. Klassenstufe mit der visuellen Programmierumgebung Scratch agile Methoden eingeführt werden. Dem Vorgehen wird

[1] Universität Stuttgart, Abteilung Berufspädagogik mit Schwerpunkt Technikdidaktik (BPT), Azenbergstraße 12, Stuttgart 70174, tobias.bahr@ife.uni-stuttgart.de
[2] Universität Stuttgart, Abteilung Berufspädagogik mit Schwerpunkt Technikdidaktik (BPT), Azenbergstraße 12, Stuttgart 70174, bernd.zinn@ife.uni-stuttgart.de

doi:10.18420/inf2022_89

unterstellt, dass das niedrigschwellige Einstiegsniveau von Scratch [Ma10] den Vorteil für die Schüler*innen und der Lehrperson bietet, den Fokus auf den methodischen Unterrichtsablauf im Kontext des agilen Projektmanagements zu richten.

Auf agile Methoden kann in späteren Klassenstufen und bei komplexeren Projektarbeiten immer wieder zurückgegriffen werden, wenn die grundlegenden Vorgehensweisen der Methode den Lehrenden und Lernenden bekannt sind. Auch die Nutzung von einzelnen agilen Methoden im Unterricht kann von den Beteiligten strukturgeben und positiv wahrgenommen werden [Br19]. Gerade bei Informatikprojekten in der Oberstufe bzw. Kursstufe oder im fächerverbindenden Profilfach Informatik, Mathematik, Physik (IMP), welches an allgemeinbildenden Gymnasien in Baden-Württemberg unterrichtet wird und bei denen die Lerninhalte spiralcurricular zunehmend komplexer werden und sich Projektarbeiten über mehrere Wochen erstrecken, bietet ein strukturiertes Vorgehen offensichtliche Vorteile. Positive Aspekte werden sowohl für den Lernprozess als auch für die Projektleistung der Schüler*innen gesehen [Br15]. Lehrpersonen können davon profitieren, dass sich durch das Project Board transparent darstellen lässt, welche Schüler*innen an welcher Aufgabe gearbeitet haben und wie der Projektfortschritt in den einzelnen Teams ist. Zudem kann es die Lehrpersonen im Hinblick auf eine transparente Leistungsbewertung der Schüler*innen unterstützen.

2 Der agile Prozess

Da an anderen Stellen der agile Prozess bereits ausführlich beschrieben und dokumentiert ist [Br19], wird der grundlegende Prozess hier nur kurz erklärt. Für weitergehende Beschreibungen und Erläuterungen wird auf das Lehrbuch *Agile Schule* verwiesen, auf das sich die hier dargestellten Prozessabläufe ebenfalls beziehen [Br19]. Schwerpunkt des vorliegenden Beitrags soll daher vor allem das umgesetzte Softwareprojekt in Scratch und die Umsetzung der Unterrichtseinheit sein. Das Pflichtenheft für die Software sieht u.a. vor, dass das Minigolf-Spiel von einem User (Spieler*in) auf drei Bahnen gespielt werden kann. Im Hauptmenü hat der User die Möglichkeit sich einen Spielernamen zu geben. Am Ende des Spiels nach der letzten Bahn, oder im Hauptmenü wählbar, soll der Highscore angezeigt werden. Mehrere User sollen das Spiel nacheinander nutzen können.

Nachfolgend wird auf die Unterrichtseinheit und dessen Ablauf eingegangen. Zu Beginn des Minigolf-Projektes wird mit einem fragend-entwickelnden Unterricht eine Anwendungsfalldiagramm erstellt (vgl. Abbildung 1), bei dem möglichst alle Anwendungsfälle des Users für den Spielverlauf visualisiert werden. Sollten die Schüler*innen bereits in Teams gut zusammengearbeitet haben, so bietet es sich an, dass die Erstellung des Anwendungsfalldiagrammes in den einzelnen Teams arbeitsgleich vorgenommen wird.

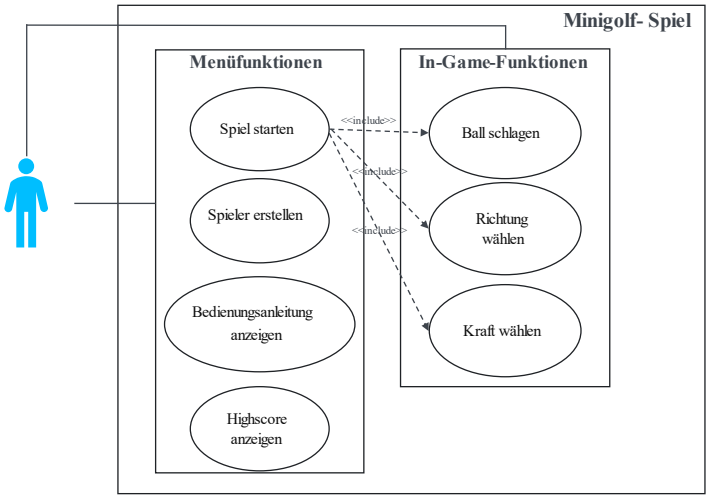


Abbildung 1: Anwendungsfalldiagramm (Use-Case-Diagramm)

Anschließend daran formulieren die Teams in Einzelarbeit oder teamübergreifend die einzelnen sogenannten Tasks[3], womit sie dann den Backlog ihres Project Boards (vgl. Abbildung 2) füllen. Tasks sind einer User Story[4] zugeordnet. Jedes Team spielt danach Planning Poker[5] um die User Stories und Tasks zu priorisieren.

[3] Tasks sind kleinstmögliche Aufgaben, die nicht mehr in Teilaufgaben zerlegbar sind. Tasks werden von einzelnen Teammitglieder bearbeitet [vgl. BR19].

[4] Eine User Story hat immer einen Namen und beschreibt in einem Satz was der Nutzer, „as a“ der Software machen möchte „I want to“, um etwas zu bekommen „so that“ [vgl. BR19].

[5] Beim Planning Poker werden Karten mit einer Einschätzung des Aufwandes von den Teammitgliedern ausgespielt. Anschließend wird eine User Story mit einem Aufwand bewertet [BR19].

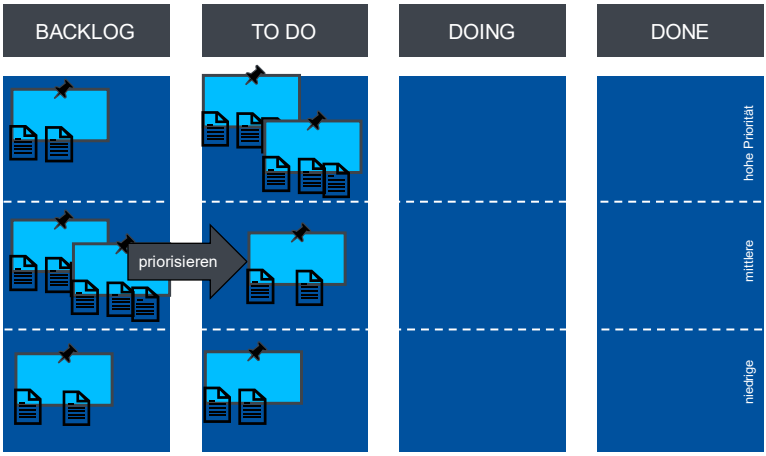


Abbildung 2: Project Board mit Backlog (vgl. Abb. 6 aus [Ka15], Abbildung 3.4 aus [Br19])

In den nächsten Unterrichtsstunden ist der Ablauf fast immer identisch. Zu Beginn finden sich die Teams in Stand-Up-Meetings vor dem Project Board zusammen, besprechen kurz welche Tasks erledigt wurden, welche als nächstes bearbeitet werden und ob Probleme aufgetreten sind. Die restliche Zeit der Unterrichtsstunde wird zur inhaltlichen Bearbeitung der einzelnen Tasks genutzt.

Reflexionsphasen stellen ein zentrales Element des agilen Prozesses dar. Daher wird am Ende der Stunde innerhalb des einzelnen Teams kurz besprochen (1.) was gut lief, (2.) was optimiert werden kann und (3.) was das Team im nächsten Schritt konkret verändern will [Br19]. Das agile Prozessvorgehen sieht auch vor, dass während der Unterrichtsstunde, bedarfs- und anlassbezogen, Stand-Up-Meetings eingefordert werden können. Somit wird ermöglicht, dass sich die Teams direkt und kurzfristig besprechen und sich über im Projektprozess aufgetreten Probleme unmittelbar und zeitnah austauschen können.

Ein weiterer wichtiger Teil der Softwareentwicklung ist das Testen der produzierten Software. Zur methodischen Umsetzung dessen ist es beispielsweise möglich, dass sich Teams gegenseitig und orientiert am Anfang aufgestellten Pflichtenheft testen. So können die Schüler*innen untereinander einen kritisch-konstruktiven Austausch üben. Basis des Feedbacks kann ein Fragebogen zu den einzelnen Funktionalitäten des Minigolf-Spiels sein. Optional steht den Schüler*innen ein didaktisch reduzierter Fragebogen, der in Usabilitystudien für Softwareprodukte eingesetzt wird, für die Testung und ggf. Bewertung zur Verfügung. Hierdurch können die Schüler*innen auch vertiefte Einblicke in die Entwicklung und Forschung im Bezugsfeld der Softwareentwicklung gewinnen. Auch hier ist zu beachten, dass den Schüler*innen die Kriterien des Usability-Fragebogens bereits zu Projektbeginn bekannt sein müssen. Andernfalls ist das Feedback zur Usability von der

Lehrperson in den entsprechenden Kontext zu setzen, da die Evaluationskriterien vorher bekannt und transparent gemacht werden müssen.

Nachdem bei den Programmen der jeweiligen Teams die Basisfunktionen vorhanden sind, findet ein Steakholdergespräch mit der Lehrperson, die weitere leistungsdifferenzierende Anforderungen an die Software stellen kann, statt. Das Minigolf-Spiel kann beispielsweise um einen Multiplayermodus für mindestens drei Spieler*innen erweitert werden. Dabei spielen die Spieler*innen immer nacheinander eine Bahn. Zusätzlich können je nach den zu erwerbenden inhaltsbezogenen Kompetenzen in der jeweiligen Klassenstufe weitere Elemente hinzugefügt werden. Beispielsweise können die inhaltsbezogenen Kompetenzen im Bereich Algorithmen im Informatik-Teil von IMP zusätzlich im Projekt fokussiert werden [Au18]. Die drei Bahnen müssen dann zusätzlich pseudozufällig ausgewählt werden, sodass sich die Reihenfolge beim Spielen beliebig ändern kann. Die eben genannten Anforderungen implizieren weitere Änderungen bzw. Erweiterungen der bisherigen Funktionen. Durch dieses Vorgehen lernen die Schüler*innen einzelne Elemente des Software Development Life Cycle (SDLC) und erhalten einen Einblick in den beruflichen Alltag einer/s Programmentwickler*in (vgl. Abbildung 3). Hierbei wurde die vereinfachte Darstellung des SDLC mit fünf Phasen gewählt. Die Darstellung gibt einen Überblick zur Unterrichtseinheit, die sich über 13 Doppelstunden erstreckt und die Entwicklung des Softwareproduktes allgemein visualisiert. Beim agilen Vorgehen wird dieser Zyklus mehrmals iteriert. Als letzten Punkt dieser Kurzbeschreibung ist bei der Durchführung das möglicherweise wichtigste Kriterium für die Umsetzung von agilen Methoden im Unterricht zu nennen, nämlich die konsequente Durchführung des Projekts. Das Einarbeiten in agile Methoden erfordert ein wenig Unterrichtszeit bei der Einführung und der Vorbereitung des Projekts. Wenn es in den ersten Wochen, oder später im Softwareprojekt zu möglicherweise unerwünschten Situationen kommt und die agile Umsetzung von den Schüler*innen in Frage gestellt wird, sollte die Umsetzung trotzdem bis zum Ende durchgeführt werden. Andernfalls können Probleme bei der klassischen Umsetzung von Softwareprojekten, wie ein nicht lauffähiges Programm, hoher Arbeitsaufwand bei den Teams und Frustration bei einzelnen Schüler*innen auftreten [Br15, Ke15].

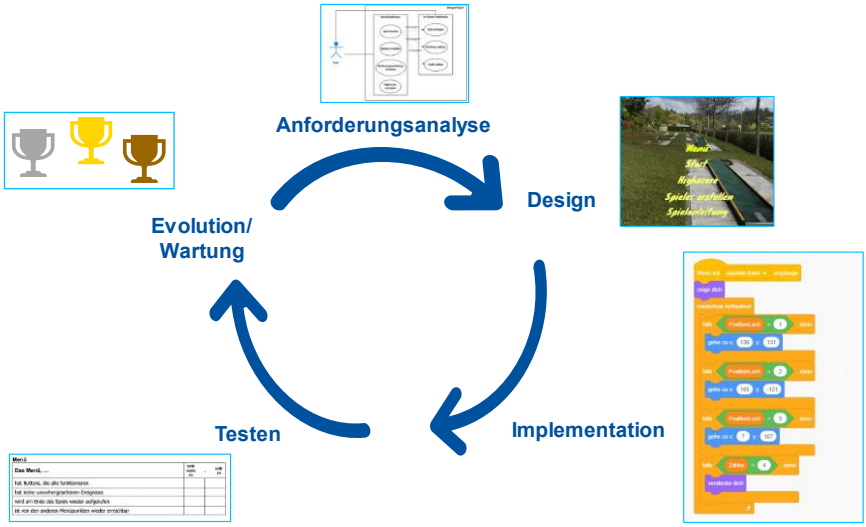


Abbildung 3: Software Development Life Cycle des Minigolf-Spiels (vgl. [Al19])

3 Kreativität im agilen Informatikunterricht

Die Erstellung eines Softwareprodukts wird vielfach als eine der Voraussetzungen für kreativen Informatikunterricht genannt [Ro08] und kann zugleich eine kompetenzförderliche Plattform, sowohl für fachliche als auch überfachliche Kompetenzen bieten. Das Commitment [Br19] der Teammitglieder am „eigenen“ Projekt, im selbstgewählten Team und in einem der Lerngruppe angepassten Projektanforderungsniveau zu arbeiten erfüllt die drei universellen Grundbedürfnisse nach der Selbstbestimmungstheorie: Streben nach Kompetenz, sozialer Eingebundenheit und Autonomie kann damit motivationsfördernd wirken [De00]. Dieses Commitment kann als höher angenommen werden, wenn die Schüler*innen nicht auf eine vorgefertigte funktionale Lösung durch das Ausführen kleinschrittiger Arbeitsaufträge hinarbeiten, sondern in einem kreativen (agilen) Prozess eigene Ideen einbringen können. Durch Differenzierungsmöglichkeiten und Offenheit in späteren Aufgabenstellungen ist es den Schüler*innen beim Minigolf-Spiel als Softwareprojekt möglich, die Aufgaben aktiv mitzugestalten. Ob sie z. B. Sounds, Wind, Hindernisse oder andere Spielelemente und -funktionen hinzufügen, können sie selbstständig entscheiden. Um neue Ideen für weitere Funktionen des Programms zu generieren, können zusätzlich Kreativitätstechniken wie Brainstorming, Mindmaps oder die Alphabet-Technik Eingang in den agilen Projektunterricht finden. Den Schüler*innen keinen Lösungsweg oder Muster

vorzugeben, ist damit eine zentrale Anforderung in der Aufgabenstellungen und soll das kreative Arbeiten von Schüler*innen sowie eigene Freiräume für Kreativität und Agilität ermöglichen [Ro08]. Als letzten Punkt sind die Gestaltungsmöglichkeiten für Schüler*innen zu nennen. Bis auf die Limitationen durch Anwendungsfalldiagramme und den curricularen Vorgaben (z. B. Bildungsplan IMP [Au18]) zu den inhaltsbezogenen Kompetenzen, sowie der vorgegebenen Programmierumgebung Scratch, können die Schüler*innen das Minigolf-Spiel, einschließlich Menübedienung, individuell gestalten.

Agile Methoden können in vielfältigen informationstechnischen Projekten und fächerübergreifenden Kontexten genutzt werden. Aspekte des kreativen Informatikunterrichts wie z. B. die Ideenanregung durch komplexe Probleme oder ein Anknüpfen an die (über)fachliche Erfahrungswelt der Schüler*innen [Ro08] kann in eine agile Umsetzung eingebunden werden. Die offene und trotzdem, durch die einzelnen agilen Methoden gut strukturierte, Bearbeitung des Projekts kann, so die Annahme, die Kreativität der Schüler*innen und ihre grundlegenden Einstellungen im Informatikunterricht wünschenswert fördern. Wünschenswert dahingehend, dass sie arbeitsfähige Strukturen, die u.a. auch in Unternehmen umgesetzt werden, lernen, als auch die konstruktiven und innovativen Aspekte der Gestaltung der Informatik applizieren können.

Heterogene Produktergebnisse sind beim methodischen Einsatz agiler Methoden in Softwareprojekten eine entscheidende Facette. Beispielsweise entscheidet sich jedes Team individuell für die Visualisierung der einzelnen Hintergründe und Kostüme. Die Lehrperson hat die Möglichkeit das Anwendungsfalldiagramm in einer Plenumsphase mit den Schüler*innen zusammen zu erstellen, oder bei entsprechendem Fachwissen und prozessbezogenen Kompetenzen, die Erstellung des Diagramms den Schüler*innen zu überlassen. Für den letztgenannten Fall ermöglicht es die Transparenz des Project Boards aus der Perspektive der Lehrperson, sicherzustellen, dass durch das Projekt alle fachlichen Kompetenzen abgedeckt werden. Das weitere Vorgehen, wie die Formulierung, das Bewerten bzw. Einstufen, das Abarbeiten, Abändern, oder Hinzufügen der einzelnen Tasks nach der Besprechung des Anwendungsfalldiagramms befähigt das Team ihr Softwareprodukt individuell zu gestalten. Zusammengefasst mit einem Satz aus dem agilen Schulmanifest: „Kreativität und Reagieren auf Gelerntes sind wichtiger als das Befolgen eines starren Plans“ [Br18]. Wichtig beim Einstufen der einzelnen Tasks ist damit, dass in der Unterrichtsverlaufsplanung ausreichend Zeit für Kreativität bei der Gestaltung eingeplant wird.

Der Projektunterricht mit agilen Methoden ist schüler- und nicht lehrerorientiert ausgerichtet. Die Lehrperson fungiert, wie bei den Anforderungen an die Lehrperson für kreativen Informatikunterricht als Coach oder Moderator [Ro08]. Im Vordergrund der Lehrperson stehen, neben dem üblichen Classroom-Management, das bedarfsgerechte Helfen und Unterstützen bei größeren anhaltenden Problemen und ggf. das Motivieren und Inspirieren der Schüler*innen. Um einer möglichen Leistungsheterogenität innerhalb der Klasse gerecht zu werden, bietet die agile Projektmethode damit generell gute Voraussetzungen, um die Individualisierung im Unterricht zu fördern. Strukturgebende Aufgaben der Lehrperson sind zudem die agile Umsetzung, das Monitoring der Arbeit und die Ak-

tualisierung der Project Boards. Steakholdergespräche, oder Vorträge, die zusätzlich eingefordert werden können. Außerdem kann die Lehrperson an Stand-Up-Meetings und Reflexionen als Beobachtende oder im Extremfall intervenierende Person (z.B. durch das Hinweisen auf bestimmte agile Werte wie bspw. Respekt) teilnehmen, um Eindrücke von der Arbeit im einzelnen Team zu bekommen.

4 Ziele des Workshops

Die Ziele des bereits mehrfach in Baden-Württemberg umgesetzten Workshops für Lehrpersonen im Profilfach IMP sind, dass die Teilnehmenden das agile Schulmanifest [Br19], Werte in agilen Projekten, den gesamten agilen Prozess, sowie die verschiedenen agilen Methoden kennenlernen [ebd.]. Darauf aufbauend sollen die Teilnehmenden ein handlungsbezogenes Wissen zum Zyklus des agilen Prozesses in reduzierter Form in der Schülerperspektive erwerben. Anschließend daran, werden die Vor- und Nachteile beim Einsatz von agilen Methoden und ihrer digitalen Umsetzung im Unterricht und den individuellen schulischen Voraussetzungen der Teilnehmenden diskutiert und reflektiert.

5 Pilotierung und Erkenntnisse

Im Rahmen einer ersten Pilotierungsstudie im Jahr 2021 nahmen N = 37 Informatik-Lehrpersonen teil. Die Lehrerweiterbildung fokussierte *Agile Methoden* und *digitale Tools zur unterrichtlichen Umsetzung*. In einer Vorabumfrage zur Weiterbildungsmaßnahme mit dem Ziel, die Weiterbildung eng an den individuellen fachlichen Voraussetzungen und Interessen der Lehrpersonen zu orientieren, gaben 85 % der Befragten an, zuvor nichts über Agile Schule bzw. agile Unterrichtsmethoden gehört zu haben. Auch wenn die Befragungsergebnisse nach der Lehrerweiterbildung auf Selbsteinschätzungen der Teilnehmenden beruhen und mögliche Verzerrungen (z. B. soziale Erwünschtheit) beinhalten, so stimmen diese positiv im Hinblick auf einen tatsächlichen Einsatz der Methode im Informatikunterricht durch die Schulungsteilnehmer*innen. Demnach gaben 80,8% der Befragten haben an, dass sie sich zutrauen, die Unterrichtseinheit Minigolf-Spiel mit Scratch durchzuführen und dabei agile Methoden anzuwenden. 65,4% haben in der Weiterbildungseinheit, entsprechend ihren Angaben, viel Neues im Bereich Agile Schule dazu gelernt (vgl. Abbildung 4). Die Tatsache, dass 85 % der befragten Lehrpersonen vor der Weiterbildung angeben, die Methode bisher nicht im Unterricht verwendet zu haben, zeigt zudem, dass agile Methoden noch keinen flächendeckenden Einzug in den IMP Unterricht an Gymnasien in Baden-Württemberg gehalten haben.

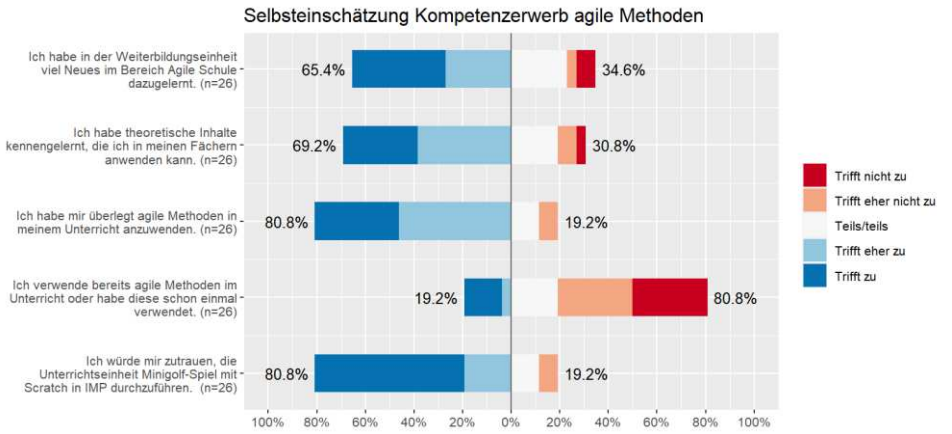


Abbildung 4: Auszug zu den Evaluationsergebnissen der Weiterbildungseinheit „Agile Schule“

6 Fazit und Limitationen

Zusammengefasst bietet der Einsatz von agilen Methoden im Informatikunterricht das Potential zur Bereitstellung adaptiver Lehr- und Lernmöglichkeiten, um im Informatikunterricht der Heterogenität der Schülerschaft gerecht zu werden und fokussiert schülerorientiertes Lernen zu unterstützen, sowie Tandem-Programmierung und individualisiertes Lernen zu adressieren. Durch das Prototyping, ein in Teilen lauffähiges Produkt zu erhalten und dieses in einem inkrementellen Prozess stetig weiterzuentwickeln, eine positive Fehlerkultur im Unterricht zu pflegen sowie fachübergreifende Kompetenzen wie Kommunikationsfähigkeit, Selbstorganisation und Kooperation prozessbezogen zu fördern, sind weitere mögliche Vorteile. Weitere Chancen bieten die hohe Unterrichtstransparenz durch das Project Board, die der Lehrperson auch eine angemessene Bewertung einzelner Schüler*innenleistungen ermöglicht. Eine Fokussierung auf agile Werte, wie Commitment, Respekt, Fokussierung und strukturiertes Feedback können förderlich für den Unterrichtsverlauf sein, indem typische Probleme im Rahmen von Gruppenarbeiten, wie beispielsweise soziales Faulenzen, Trittbrettfahren oder Gimbel Effekte geringgehalten werden [Ha75]. Wie Eingangs bereits berichtet, haben agile Methoden ihren Ursprung in unternehmerischen und wirtschaftlichen Kontexten. Vor dem Hintergrund der Digitalisierung der Arbeits- und Berufswelt [Zi17] sowie der Forderung nach Agilität von Personen, Organisationen, Strukturen und Prozessen in ebendieser Arbeits- und Berufswelt, erscheint es wünschenswert entsprechend hierfür notwendige Kompetenzen bereits in der Schule zu fördern.

Flankiert mit den Erfahrungen aus Praxisberichten [Br15, Br19, Ka15, Gö11, Ke15] ist der Einsatz von agilen Methoden kein Patentrezept für einen guten Projektunterricht, kann

jedoch dessen Umsetzung, insbesondere durch seine strukturgebende und gleichzeitig flexible anpassungsfähige Methodik, unterstützen. Die Evidenz zu den tatsächlichen Wirkungseffekten des methodischen Vorgehens im Unterricht und Kontext der agilen Projektmethode im Hinblick auf die Schüler*innen, ist nach den uns vorliegenden Daten gering und umfasst mehrere Forschungsdesiderata. Sowohl im Hinblick auf die kognitiven, motivationalen als auch affektiven Merkmale von Lernenden, wären im Bezugsfeld Interventionsstudien interessant, um evidenzbasierte Aussagen zur tatsächlichen Wirkung treffen zu können.

Literatur

[Al19] Alpar, P., Alt, R., Bensberg, F., Weimann, P. (2019). Phasenmodelle in der Softwareentwicklung. In: Anwendungsorientierte Wirtschaftsinformatik. Springer Vieweg, Wiesbaden. https://doi.org/10.1007/978-3-658-25581-7_13

[Au18] Autorengruppe des Ministeriums für Kultus, Jugend und Sport Baden-Württemberg (2018). Bildungsplan zum Profilfach Informatik, Mathematik, Physik (IMP). BILDUNGSPLAN DES GYMNASIUMS. http://bildungsplaene-bw.de/site/bildungsplan/get/documents/lsbw/export-pdf/depot-pdf/ALLG/BP2016BW_ALLG_GYM_IMP.pdf

[Br15] Brichzin, P. (2015). Agile Softwareentwicklung. Erfahrungsbericht eines Oberstufenprojekts im Wahlpflichtunterricht. In: Gallenbacher, J. (Hrsg.), Informatik allgemeinbildend begreifen. Bonn: Gesellschaft für Informatik e.V.. 63-72.

[Br19] Brichzin, P., Kastl, P., & Romeike, R. (2019). Agile Schule: Methoden für den Projektunterricht in der Informatik und darüber hinaus. Hep Verlag.

[Co05] Coram, M. & Bohner, S. (2005). The impact of agile methods on software project management. 12th IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS'05). 363-370.

[De00] Deci, Edward. L. & Ryan, Richard. M. (2000). The „What“ and „Why“ of Goal Pursuits: Human Needs and the Self-Determination of Behavior. In: Psychological Inquiry 11(4). 227-268.

[Di15] Dietrich, L., Gramm, A., Kastl, P. & Romeike, R. (2015). Ein Bild vom Wesen der Softwareentwicklung: Erfahrungen aus zwei agilen Projekten. In: Gallenbacher, J. (Hrsg.), Informatik allgemeinbildend begreifen. Bonn: Gesellschaft für Informatik e.V.. 83-92

[Gö11] Göttel, T., (2011). Agiler Informatikunterricht: Soziale Aspekte der professionellen Softwareentwicklung im Schulunterricht erfolgreich erfahrbar machen. In: Thomas, M. (Hrsg.), Informatik in Bildung und Beruf – INFOS 2011 – 14. GI-Fachtagung Informatik und Schule. Bonn: Gesellschaft für Informatik e.V.. (S. 37-46).

[Ha75] Hackman, J.R., & Morris, C.G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. Advances in Experimental Social Psychology, 8, 45-99.

[Ka15] Kastl, P., März, S. & Romeike, R., (2015). Agile Softwareentwicklung im Informatikunterricht. Ein Best-Practice-Beispiel am Spiel “Pengu'. In: Gallenbacher, J. (Hrsg.), Informatik allgemeinbildend begreifen. Bonn: Gesellschaft für Informatik e.V.191-200.

[Ke15] Kerber, L., Kastl, P. & Romeike, R., (2015). Agiler Informatikunterricht als Anfangsunterricht. In: Gallenbacher, J. (Hrsg.), Informatik allgemeinbildend begreifen. Bonn: Gesellschaft für Informatik e.V.. (S. 211-218).

[Ma10] Maloney, J., Resnick, M., Rusk, N., Silverman, B., & Eastmond, E.(2010). The Scratch Programming Language and Environment. ACM Trans. Comput. Educ. 10(4), 15 pages. https://doi.org/10.1145/1868358.1868363

[Ri09] Rico, D. F. & Sayani H. H. (2009). Use of Agile Methods in Software Engineering Education. 2009 Agile Conference.174-179.

[Ro08] Romeike, R. (2008). Kreativität im Informatikunterricht. Potsdam, Universität Potsdam (Doctoral dissertation, Dissertation).

[Sc13] Schumacher, C., Rengstorf, F., & Thomas, C. (Eds.). (2013). Projekt: Unterricht: Projektunterricht und Professionalisierung in Lehrerbildung und Schulpraxis. Vandenhoeck & Ruprecht.

[Tr11] Traub, S. (2011). Selbstgesteuert lernen im Projekt? Anspruch an Projektunterricht und dessen Bewertung aus Sicht von Lehrenden und Lernenden. Zeitschrift für Pädagogik, 57 (1), 93–113.

[Wa09] Wasmann-Frahm, A. (2009). Kompetenzentwicklung durch Projektunterricht. Unterrichtswissenschaft, 37 (1), 77-96.

[Wi03] Williams, L. & Cockburn, A. (2003). Agile software development: it's about feedback and change. in Computer, 36 (6), 39-43.

[Zi17] Zinn, B. (2017). Digitalisierung der Arbeit – Kompetenzerwartungen des Beschäftigungssystems und didaktische Implikationen. In: Bonz, B., Schanz, H. & Seifried, J. (Hrsg.): Berufsbildung vor neuen Herausforderungen – Wandel von Arbeit und Wirtschaft, Berufsbildung konkret, Band 13, Baltmannsweiler: Schneider Verlag Hohengeheren GmbH, S. 163-176.

Geodaten als open data für die Künstliche Intelligenz

Geodaten als Open Data für die Künstliche Intelligenz

Anna Hoffmann[1] Reinhard Kahle[2]

Die moderne Künstliche Intelligenz lebt von Daten. Ein hoher Datenschutzstandard erschwert zuweilen den Zugang zu Daten. Allerdings stehen aufgrund europäischer Initiativen inzwischen durch „Copernicus“ und „Galileo“ umfassende Geodaten als open data zur Verfügung. Auf diesem Workshop diskutieren wir die Möglichkeiten und Herausforderungen, die sich durch die Verwendbarkeit dieser Daten mittels Künstlicher Intelligenz ergeben. Das gilt einerseits für die technischen Anforderungen an Daten und Software; andererseits betrachten wir neue und innovative Anwendungen, z.B. in der Umweltforschung, die der Gesellschaft zu gute kommen.

In fünf kürzeren Beiträgen wird das Thema von verschiedenen Akteuren aus unterschiedlichen Perspektiven beleuchtet.

Daran schließ sich eine längere gemeinsame Dialogrunde aller Gäste an, die im Sinne des Dialogformats nach David Bohm geführt wird. Davis Bohm war ein US-amerikanischer Quantenphysiker und Philosoph, der auch mit Einstein gearbeitet hat. Bohm hat sich bemüht, durch sein zusammen mit Martin Buber entwickeltes Dialogformat mehr zu Offenheit und Verständigung zwischen verschiedenen Standpunkten beizutragen. Im Zentrum eines Dialogs nach Bohm stehen deshalb nicht der Austausch von Argumenten, sondern das voneinander Lernen sowie das Entwickeln von weiterführenden Fragen und Lösungsoptionen.

Ziel unseres Dialogs auf diesem Workshop soll es darum auch sein, gemeinsam Fragen und Handlungsansätze zu entwickeln, die das Thema nach vorne bringen können.

Zu den Gesprächspartner gehören unter anderem:

Das Start-Up *Beetle ForTech*, die den Copernicus Masters 2021 und den Galileo Masters 2021 gewonnen haben.[3]

Das Verbundprojekt *NaLamKI* des Fraunhofer HHI in Berlin.[4] Die Schwerpunkte des HHI innerhalb von NaLamKI liegen in der Entwicklung von KI-Verfahren zu Analyse von Fernerkundungsdaten, zur Erfassung und Integration von Expertenwissen, zur Modellierung landwirtschaftlicher Prozesse sowie zur 5G-Vernetzung.

[1] Hoffmann Consulting & Facilitation, Berlin & Potsdam anna@anna-hoffmann-consulting.de
[2] Universität Tübingen, Carl Friedrich von Weizsäcker-Zentrum, Doblerstr. 33, 72074 Tübingen reinhard.kahle@uni-tuebingen.de
[3] https://beetlefortech.com/
[4] https://www.hhi.fraunhofer.de/abteilungen/vit/projekte/nalamki.html und https://nalamki.de/ziele/

doi:10.18420/inf2022_90

Die Moderation und Organisation des Workshops liegt in den Händen von Anna Hoffmann und Reinhard Kahle.

QSAP - Quantum Software and Applications

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1085

High-Performance and Quantum Computing for Students

Peter Faber[1] Helena Liebelt[2]

Abstract: Many technical advances have been applied in the area High-Performance Computing for the first time, and this also seems to hold for Quantum Computing. Traditionally, however, experts in these fields come from a diverse educational background. Recently, Technische Hochschule Deggendorf (Deggendorf Institute of Technology, DIT) introduced a new master's program „High-Performance Computing / Quantum Computing" to offer students a direct way into the application specific aspects of these field of expertise. To our knowledge, this is the first course of studies of its kind. This work describes this new program of studies and tentatively identifies some of the first lessons learned so far.

Keywords: Master's program; High-Performance Computing; Quantum Computing

1 Introduction

The field of both, High-Performance Computing (HPC) and Quantum Computing (QC), is becoming more and more important in today's world that relies on large scale data processing and ever more complex problems to be solved.

Thus, it should come as no surprise that markets in this area show rising numbers, even through pandemic times, and are gaining public interest, both locally in Germany as well as globally [Fö22; In21; Ka22; Ru21].

Therefore, Technische Hochschule Deggendorf (Deggendorf Institute of Technology, DIT) has introduced a new master's program called „High-Performance Computing / Quantum Computing" (HPC/QC). To our knowledge, this is the first course of studies of its kind: Technische Hochschule Deggendorf is a so-called university for applied sciences. This type of university focuses much more on aspects of applications than typical universities. This focus on immediate applicability in computing centers also distinguishes this course of studies from similar programs like the Master in Quantum Engineering at ETH Zürich [ET], the Master Quantum Engineering at University of Saarbrücken [Un], or the joint Master's Program in Quantum Science & Technology at Technische Universiät München and Ludwig-Maximilians-Universität München [TL] that have also been launched recently and turn their focus more specifically to the foundations and technologies of QC in particular.

[1] Technische Hochschule Deggendorf (THD / DIT), Dieter-Görlitz-Platz 1, 94469 Deggendorf, Germany
peter.faber@th-deg.de
[2] Technische Hochschule Deggendorf (THD / DIT), Dieter-Görlitz-Platz 1, 94469 Deggendorf, Germany
helena.liebelt@th-deg.de

doi:10.18420/inf2022_91

In contrast, the HPC/QC program tries to give students insights into immediately available applications. This also means combining high-performance and quantum computing in one course of studies. This makes sense for multiple reasons: Firstly, as can be seen throughout the history of computing, new performance improving technologies are mainly picked up by High-Performance Computing (HPC) architectures before reaching more general computing devices. Secondly, the immediate availability of an already existing and growing job market, which the building and maintenance of HPC systems offers, establishes a secure outlook for students. However, even for purely traditional HPC workloads, the mindset needed for the development of quantum circuits represents a valuable asset. Thus, we see large overlaps in the application of these two topics of the master program and exploit resulting synergies.

2 Structure of the Program

Since bachelor programs in Germany usually extend over seven semesters (210 ECTS), master programs typically cover three semesters (180 ECTS). Most courses at THD/DIT cover 5 ECTS in order to be able to reuse modules; therefor, modules are generally tailored to contain one course with an appropriate workload.

Figure 1 supplies a brief overview of the curriculum of this course of studies, an adapted version of the presentation at the program's website [Te21]. The program aims to offer a structured approach to developing expertise in design, maintenance, and development for modern HPC systems, as well as the newly appearing QC systems.

As can be gleaned from Figure 1, the overall course can be divided into four pillars:

1. Hardware Design and Efficiency
2. Software Engineering and Optimization
3. System Design and Application
4. General Competences

In the hardware design part, students in their first semester are introduced to (quantum) physics as a foundation of QC. This course can also be seen as an introduction to scientific reasoning and thus builds a base for the whole course of studies.[3] Students are also introduced to technologies fundamental to both HPC and QC systems; e.g., an understanding of networking technology is important not only for the operation of classical HPC systems but also for estimating problems and problem sizes that may be approached with QC.

The second pillar aims at introducing the students to programming techniques and frameworks for both, HPC as well as QC architectures. Students are not only introduced to parallel

[3] Since the course of studies actually started in the winter semester, this basic course was offered in both, the winter and the summer semester at the start of the program.

Master Program HPC/QC (M.Sc.)		Hours per week (SWS)					
Module / course	**Contents**	**SWS**	**1. Sem.**	**2. Sem.**	**3. Sem.**	**ECTS**	**Exam**
Hardware Design and Efficiency							
Physics for HPC/QC	Physical basics (in particular for QC)	4	4			5	written
Computer Architectures for HPC/QC	Concepts of modern computing systems (cluster systems, quantum computers)	4		4		5	written
Networks for HPC/QC	General and advanced topics of networking (advanced technologies)	4		4		5	written
Software Engineering and Optimization							
Software Engineering	Advanced methods of software engineering	4	4			5	written
HPC/QC Programming Lab	Modern methods of HPC and QC programming (frameworks and concepts)	4	4			5	project
Optimization Methods	Optimizing compilers, performance analysis & optimizations	4		4		5	project
System Design and Application							
HPC/QC Technology	Technical particularities of HPC and/or QC systems	4	4			5	project
HPC/QC Infrastructure	Infrastructure of modern computing systems in computing centers	4		4		5	written
System Design and Application of HPC/QC Systems	Complex system design (realizing an actual project from scratch)	4		4		5	project
General Competences							
Advanced Mathematics for HPC/QC	Mathematical foundations of scientific analyses	4	4			5	written
Advanced Mathematics and Physics for HPC/QC	Advanced mathematical and physical topics (scientific calculations)	4		4		5	written
Faculty Elective I**	Elective, in particular for deepening knowledge relevant for the area of HPC/QC	4	4			5	depends on subject
Faculty Elective II**	Elective, in particular relevant for the student's area of expertise	4			4	5	depends on subject
Master's colloquium	Presentation and disputation of the master's thesis	2			2	2	oral
Master's thesis	Academic research and writing	0			0	23	thesis
Total SWS		54 SWS	24	24	6		
Total ECTS			30	30	30	90 ECTS	

Fig. 1: Curriculum of the HPC/QC master program

programming using well known frameworks such as OpenMP [Op21] and MPI [Me21] (and also exposed to modern optimizing compilers etc.), but also to frameworks for quantum computing. In particular, simulator environments, partly based on HPC technology, are introduced for this purpose, including (but not restricted to) the Intel Quantum Simulator (IQS) [Gu20] and Google's cirq [Ci21].

The system design part leads students through maintenance and development tasks for HPC and/or QC systems. Again, HPC and QC systems are usually viewed together here, with HPC systems building frontends for QC systems. The final goal in this field of expertise is to develop a complete system from scratch during the second semester; a task for which students can finally draw from their complete knowledge attained in the preceding courses.

The field of general competencies contains a special approach to electives selection: Experts in the areas of HPC and QC oftentimes come from a variety of different careers and academic backgrounds, such as computer science, but also mathematics, physics, electrical engineering etc. Therefore, elective courses during the first semester should serve to deepen the student's knowledge in areas that are relevant to HPC or QC, in particular if the individual student's prior work and experiences themselves may not have provided adequate deeper knowledge in that area. This helps to ascertain a common level of knowledge for all students. However, it also implies that study program coordinators evaluate the applicability of different available courses for each student separately.

As a university of applied sciences, THD/DIT strives to build its programs in collaboration with other research facilities and industry partners alike to maximise the student's exposure to new developments and technologies. The partner network includes most of the major market players in both QC and HPC industries. Thus, courses are developed holistically, regularly receiving presentations and multi-day workshops from partners. Established regular „office-hours" with industry partners act as a direct line between students and the engineering teams from Intel, Microsoft and others to provide unbureaucratic support and guidance. Furthermore students have the opportunity to engage with industry mentors from various segments. Contributions by industry partners are planned together with the partners so that, depending on the possibilities, different kinds of contributions (from single presentations to multi-day workshops) can be integrated into the corresponding courses.

3 Experiences

The master program, primarily addressing students with degrees in mathematics, physics and computer science, is available for students from anywhere around the globe. Therefore, applicant numbers tend to be quite high: since contents and quality standards in education across countries and universities vary significantly, the application process for this program had to include an efficient way to identify suitable applicants. To that end, applicants have to pass a partially automated application test that contains questions from the fields of mathematics / logic, physics, operating systems and networks, programming, and modelling.

The test consists of a mixture of different question types ranging from multiple choice questions, through calculations to descriptions of their work in their previous studies. These tests were processed in one semester using a quiz in a separately managed instance of the school's e-learning platform Moodle, and in one semester using ARLA [El20], an experimental, automatic test system that is being developed at European Campus Rottal Inn (ECRI, part of THD). Both approaches proved valid alternatives for this task.

In its first semester, the program attracted 631 applications. 25 applicants were admitted to the program. In the summer semester (the regular start), 453 applications were received, 23 students were admitted. 26 students are currently registered. Thus, the rigorous test regime effectively helped to filter out suitable candidates and remove bogus applications.

The student's performance so far can be seen as adequate: On the German scale from 1 (best) to 5 (failure), students in their first semester reached an average of about 2.4.

On the other hand, students seem to be content with their course of studies. Although feedback up to now does not yet allow for statistical soundness, the general attitude towards the program according to a feedback loop of six anonymous students (and an anonymous follow-up interview) was quite positive. On a scale from 1 (not at all) to 5 (completely), students said on average:

1. they feel well-prepared for their future job: 4.0
2. they feel the subjects build a good fit together: 4.2
3. they recommend this course of studies: 4.3
4. the course of studies matched their expectations: 4.3
5. they were pleased with the invitations of guest speakers from the industry: 4.7

Students also named points for improvement such as better module descriptions online and that the mode of exams during the pandemic opened too many possibilities for cheating, which will be addressed by the coordinators of the program.

4 Conclusions

The new master program „High-Performance-Computing / Quantum Computing“ has taken off with a promising start. The application process seems to have successfully picked committed students. Since the program has only just begun, it is too early for a reliable quantitative analysis yet, but all in all, the program seems to be on the right track.

Further impulses are to be expected from an accreditation process that is currently running as well as from the new bachelor's program „Data Center Management“ [Te22] that is currently being introduced and should complement this program on the bachelor level.

References

[Ci21] Cirq Developers: Cirq, version v0.12.0, See full list of authors on Github: https://github .com/quantumlib/Cirq/graphs/ contributors, Aug. 2021, URL: https://doi.org/10.5281/zenodo.5182845.

[El20] Elkout, A.: Developing a Web Portal for Academic Readiness Level Assessment, MA thesis, Technische Hochschule Deggendorf / Deggendorf Institute of Technology, 2020.

[ET] ETH Zürich: Master in Quantum Engineering, last accesses 2022-07-10, URL: https://master-qe.ethz.ch/.

[Fö22] Förster, M.: Mainframe-Ausbildung: Glänzende Berufsaussichten, fehlt jedoch im Studium meist, Text in German, 2022, URL: https://www.heise.de/news/Mainframe-Ausbildung-Glaenzende-Berufsaussichten-fehlt-jedoch-im-Studium-meist-7101451.html, visited on: 05/25/2022.

[Gu20] Guerreschi, G. G.; Hogaboam, J.; Baruffa, F.; Sawaya, N. P.: Intel Quantum Simulator: A cloud-ready high-performance simulator of quantum circuits. Quantum Science and Technology 5/3, p. 034007, 2020.

[In21] Intersect360: Total HPC Market Revenue Grew 5.2% to $41.0 Billion in 2021, Says Intersect360 Research, 2021, URL: https://www.intersect360.com/total-hpc-market-revenue-grew-5-2-to-41-0-billion-in-2021-says-intersect360-research/, visited on: 05/25/2022.

[Ka22] Kaufmann, W.: High Performance Computing, Modularität und Made in Germany, Text in German, 2022, visited on: 05/25/2022.

[Me21] Message Passing Interface Forum: MPI: A Message-Passing Interface Standard Version 4.0, June 2021, URL: https://www.mpi-forum.org/docs/mpi-4.0/mpi40-report.pdf.

[Op21] OpenMP Architecture Review Board: OpenMP Application Program Interface Version 5.2, Nov. 2021, URL: https://www.openmp.org/wp-content/uploads/OpenMP-API-Specification-5-2.pdf.

[Ru21] Russel, J.: The (Resilient) HPC Market Writ Large in 100-plus Slides by Hyperion, 2021, URL: https://www.hpcwire.com/2021/06/28/the-resilient-hpc-market-writ-large-in-100-plus-slides-by-hyperion/, visited on: 05/25/2022.

[Te21] Technische Hochschule Deggendorf / Deggendorf Institute of Technology: HPC/QC master program website, 2021, URL: https://www.th-deg.de/qc-m, visited on: 05/26/2022.

[Te22] Technische Hochschule Deggendorf / Deggendorf Institute of Technology: Data Center Management bachelor program website, 2022, URL: https://www.th-deg.de/dcm-b, visited on: 05/26/2022.

[TL] Technische Universiät München; Ludwig-Maximilians-Universität München: Master's Program in Quantum Science & Technology, last accesses 2022-07-10, URL: https://www.mcqst.de/support/master-in-qst/.

[Un] Universiät des Saarlandes: Master Quantum Engineering, last accesses 2022-07-10, URL: https://www.uni-saarland.de/studium/angebot/master/quantum-engineering.html.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1093

Implementation Aspects for Linking Data in Hybrid Quantum Applications

Markus Zajac[1] Michelle Dang[2] Uta Störl[3]

Abstract: Linking logically related data for applications in production and logistics remains a challenge. Links between logically related data are difficult to establish at the time of collection. Moreover, executing searches for the right data often requires a full collection scan. In this short paper, we address implementation aspects of an already designed hybrid system that uses a quantum search algorithm for linking data.

Keywords: Hybrid polystore systems; Quantum computing applications; Data linkage; Quantum search; QRAM

1 Introduction

Managing and analyzing production and logistics data remains a challenge. One of these challenges concerns the interconnection of data [MSN17, Wa18]. Data, in particular machine and sensor data, data from production and logistics processes, product and parts data and manufacturing orders is collected and stored in a decentralized manner (more precisely in the different types of database systems). This causes a loss of connection to other stored data records, although there is a logical connection between them. However, when implementing industrial use cases, it is important to link logically related data, so this circumstance must be mastered first. This can lead to simple data linkage search queries having a worst case complexity of $O(n)$ each time (assuming n records have to be searched). Search queries can be significantly accelerated by secondary indexes. However, using secondary indexes for faster data access to distributed data is not always trivial and can have disadvantages. The various data-driven services of the use cases with their individually tailored data views, require the use of many different secondary indexes in a distributed data storage. In addition to this organizational aspect, maintaining secondary indexes poses another problem. Index structures must be maintained during write operations while ensuring consistency between secondary indexes and data tables [QCH18]. In [ZS22], we describe a typical use case in the industry, investigate the reason for this runtime complexity, and propose a new approach to data linking. The approach forgoes the use of secondary indexes and rethinks the existing problem for linking logically related data. It conceives itself in the sense of PISQ (Perfect

[1] University of Hagen, Databases and Information Systems, Hagen, Germany markus.zajac@fernuni-hagen.de
[2] University of Hagen, Hagen, Germany michelle.dang@studium.fernuni-hagen.de
[3] University of Hagen, Databases and Information Systems, Hagen, Germany uta.stoerl@fernuni-hagen.de

doi:10.18420/inf2022_92

Intermediate-Scale Quantum computing) [BSA21]. Central components of the approach include an interface between classical and quantum data (here we specifically take up the QRAM idea) as well as Grover's quantum search algorithm. The goal is to establish connections between logically related data on demand using appropriate queries, each with sublinear time. This short paper describes some aspects of the implementation of the concept described in [ZS22]. Section 2 briefly reviews the relevant work. Section 3 addresses implementation aspects before Section 4 describes a summary and next steps.

2 Related Work

Using the principle of a quantum superposition one showed, that a database with N numeric entries ($N = 2^n$, where n bits are needed to store the numeric entries) can be searched using $\sqrt{N}$ queries [JK20]. At the beginning of the Grover's quantum search algorithm, a sufficiently large register is prepared in an equal superposition of all states

$$|\psi\rangle = \frac{1}{\sqrt{N}} \sum_{x=0}^{N-1} |x\rangle. \tag{1}$$

Then the so-called Grover iteration

$$G = (2|\psi\rangle\langle\psi| - I)O \tag{2}$$

is applied $O(\sqrt{N})$ times [NC11]. The element searched for is specified via the oracle O. The search term is to be understood as abstract: It refers to a solution space and not to search in database systems. One way to provide quantum algorithms with input data is the so-called Quantum Random Access Memory (QRAM) [KH20, LB20, BSA21]. For our work we take up the idea of QRAM. We consider the following works as important or motivating for us. Papers [We20] and [We21] present data encoding techniques for quantum computing and propose encoding for use with QRAM. Input data can also be realized via quantum circuits instead of QRAM. In [Ka21], a two-qubit search is implemented using this method. There, searching for a specific *value code* (within unsorted values) returns an index. Even though this is an example for teaching purposes, it should still be applicable in other contexts. Paper [SW19] is particularly interesting because it relates quantum-based search problems using a hybrid approach. For the realization of a Recommendation System, a database with over 12,000 entries is used. The entire relational database is however not mapped as a quantum system, but only its relevant features. These correspond to primary keys of a relational classic database. On the reduced quantum database, Quantum k-NN and Grover's algorithm are applied. In [AMA18] and [AMA19] the authors proved the effectiveness of Grover's algorithm in NoSQL databases in a theoretical study and conclude that it can significantly reduce search time compared to classic algorithms. These database systems are increasingly used for storing data in production and logistics. The embedding of quantum technologies in database architectures is also discussed in [Gr21]. There, the *SHM3P* database is presented as an alternative of an IoT database for which an embedding is envisioned.

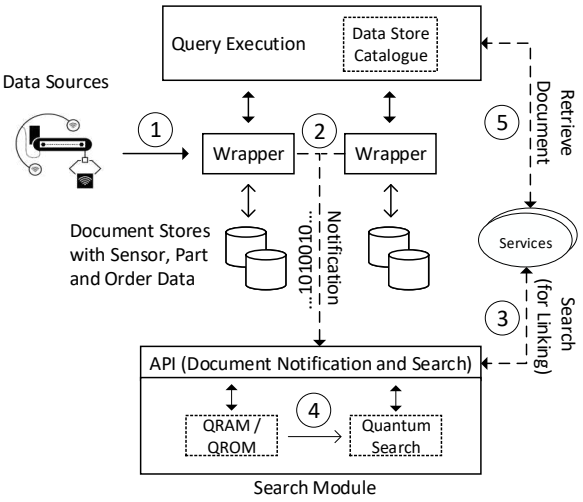


Fig. 1: Use case's hybrid polystore architecture (following [ZS22])

3 Implementation Aspects

In implementing the concept [ZS22], we have so far addressed two issues: First, data acquisition and storage in QRAM/QROM, and second, adaptations of the Grover's algorithm to exploit information from QRAM/QROM for data linking (see Figure 1). We address these issues in the following.

Test data and polystore system: For our use case, we construct two document stores (MongoDB instances) as a minimal example of a polystore system. The document store *Container* provides sensor information from shipping containers and the document store *Order* stores the orders of customers. As reference for the container data, we use an extracted and anonymized dataset which was published by the company DB Schenker under the CC BY 4.0 license. To begin with, we will focus on data from the light sensor. Sensor values are collected at different times and are stored with further information (like time, coordinates etc.) in JSON documents in document stores. To make the light sensor values accessible for quantum search, they are first encoded together with the IDs of the documents as a bit sequence, transmitted to the *Search Module*, and stored in a quantum memory (QRAM/QROM) in the order of their arrival. For a search hit, the corresponding document ID is to be delivered (see following explanations).

Quantum Random Access Memory (QRAM): QRAM can be understood as an interface for the exchange of classical and quantum data [KH20]. In QRAM multiple addresses can

be queried by querying them in superposition. This process can be expressed by means of a so-called LOAD function as follows [We21, MGM20]:

$$\frac{1}{\sqrt{k}} \sum_{i=0}^{k-1} |a\rangle_i |0\rangle \xrightarrow{LOAD} \frac{1}{\sqrt{k}} \sum_{i=0}^{k-1} |a\rangle_i |s_a\rangle \tag{3}$$

The first register (address register) is in a superposition of k queried memory addresses, the second register is the output register in superposition. Thereby $|a\rangle_i$ specifies the memory address and $|s_a\rangle$ represents the content at the i-th location. Algorithms, such as the Grover's algorithm, use the loaded data as a starting point. There are various QRAM concepts with their specific advantages and disadvantages. The authors find it rather difficult to make a general commitment to one concept. Based on a use case, the advantages and disadvantages must be weighed up. A general distinction is made between two types of QRAM: an "explicit" and an "implicit" QRAM (in the case of implicit memory, one also speaks of QROM) [MGM20]. We'll take a brief review of the two types of memory. In "explicit" QRAM, data is stored in real physical qubits. These data are read out or queried via additional circuits. The main disadvantage is the high requirement for logical qubits, which is $O(2^n)$ for n-bit addresses [MGM20]. The main advantage is the fact that the circuit of this memory type only needs to be compiled and optimized once.

In QROM, data ("database") is stored directly in a circuit itself, without the need for explicit qubits to store data. The main disadvantage is that content of the memory must be known in advance, the main advantage is that QROM circuits can be well optimized. Simple circuits with large depth and small width needs $O(n)$ qubits for n-bit addresses [MGM20]. QROM can be used in the sense of a lookup table. Today, however, there are still no efficient implementations of these concepts [LB20]. Furthermore, it can be assumed that the memory must be actively error corrected for both types. Nevertheless, the community mentioned further below offers several reference implementations. We evaluate these implementations for use with the Grover's algorithm. However, the high demand for qubits of the "explicit" QRAM didn't allow us to construct a reasonable minimal example. Therefore, we are currently evaluating QROM implementations. For the use case itself, the QROM idea is rather not optimal in the opinion of the authors, but still usable as an interface. Incoming sequences can either be buffered to generate an optimal QROM circuit ("database") at certain times. If incoming sequences should be considered directly, then a recompilation and optimization of the circuit would be necessary each time. [BSA21] also postulates that initially a "good" QROM realization is sufficient.

Grover's algorithm: Besides the type of memory, we also consider how the data should be represented in it. For the use case, we need to encode $(document\ ID, light\ value)$ tuples as bit sequences in memory. To achieve this, we represent a tuple as a bit sequence, which logically consists of two bit sequences of defined length: The first sub-sequence represents *document ID*, the second *light value*. In terms of the Grover's algorithm, bit sequences correspond to states in the solution space (there is a maximum of 2^s distinguishable states when sequences consist of s bits). Before the algorithm is executed, the sequences are

loaded into the above-mentioned register $|s_a\rangle$, which is in a superposition (according to the start condition in the Equation 1). A reference implementation of the algorithm has to be modified in such a way that the Grover oracle marks the solution (the state corresponding to a sequence) whose sub-sequence corresponds to the searched light value. In this state, both the document ID and the searched value are encoded. Based on this ID, the corresponding document can be retrieved from the polystore system mentioned above. Applications can search for different values and in this way link the logically related data on demand.

Technologies used and first experiments: We use the open source Quantum Development Kit (QDK) with the language Q# and QRAM reference implementations of *qsharp-community* on GitHub[4]. We have already started to use the IonQ quantum computer[5] for experiments with small data sets to test Grover's algorithm Oracle implementations. On the one hand, we intended to ensure that the code is not only executable on the IonQ simulator, but also on a machine. This has something to do with the fact that QPUs are more constrained (they are subject to so-called profiles), while a simulator is more tolerant. Secondly, we wanted to observe statistical deviations and tendencies of a NISQ machine. In the tests, we used sequences consisting of four qubits in order to search for specific values there. For statistical relevance, 1,000 resp. 10,000 shots were performed. We found strong deviations especially when multiple search hits are expected (i.e., when searched values are present multiple times).

4 Summary and Future Work

In this short paper, we address implementation aspects of a novel approach to fast linking logically related data on demand. In addition to interfaces such as QRAM/QROM, the Grover's algorithm is used for this purpose. At the same time, this contribution represents lessons learned. In our experience, implementation ideas or concepts and existing reference implementations do not always accompany each other smoothly. This in turn helps us to develop new ideas and suggest possible improvements. In the next steps, we continue to evaluate QROM as an interface, as well as other data encoding variants. We also develop a method to compare the asymptotic runtime of hybrid and classical systems.

Acknowledgments

This work has been funded by Deutsche Forschungsgemeinschaft (German Research Foundation) - 385808805.

[4] https://github.com/qsharp-community/qram

[5] Provided as part of the Azure Quantum Credits program

Bibliography

[AMA18] Amellal, Hicham; Meslouhi, Abdelmajid; Allati, Abderahim El: Reduce Data Processing Time in NoSQL Databases Based on Grover's Algorithm. In: Proceedings of the 3rd International Conference on Smart City Applications. SCA '18. Association for Computing Machinery, 2018.

[AMA19] Amellal, Hicham; Meslouhi, Abdelmajid; Allati, Abderahim El: Processing Unstructured Databases Using a Quantum Approach. In (Ben Ahmed, Mohamed; Boudhir, Anouar Abdelhakim; Younes, Ali, eds): Innovations in Smart Cities Applications Edition 2, Lecture Notes in Intelligent Transportation and Infrastructure, pp. 275–285. Springer International Publishing, 2019.

[BSA21] Bertels, Koen; Sarkar, Aritra; Ashraf, Imran: Quantum Computing - From NISQ to PISQ. IEEE Micro, 41(5):24–32, 2021.

[Gr21] Groppe, Sven: Semantic Hybrid Multi-Model Multi-Platform (SHM3P) Databases. In (Jain, Sarika; Groppe, Sven, eds): Proceedings of the International Semantic Intelligence Conference 2021 (ISIC 2021), New Delhi, India, February 25-27, 2021. volume 2786 of CEUR Workshop Proceedings. CEUR-WS.org, pp. 16–26, 2021.

[JK20] Jóczik, Szabolcs; Kiss, Attila: Quantum Computation and Its Effects in Database Systems. In: New Trends in Databases and Information Systems, volume 1259 of Communications in Computer and Information Science, pp. 13–23. Springer International Publishing, 2020.

[Ka21] Kain, Ben: Searching a quantum database with Grover's search algorithm. American Journal of Physics, 89(6):618–626, 2021.

[KH20] Kang, Yujin; Heo, Jun: Quantum Minimum Searching Algorithm and Circuit Implementation. In: 2020 International Conference on Information and Communication Technology Convergence (ICTC). IEEE, pp. 214–219, 2020.

[LB20] Leymann, Frank; Barzen, Johanna: The bitter truth about gate-based quantum algorithms in the NISQ era. Quantum Science and Technology, 5(4):044007, sep 2020.

[MGM20] Matteo, Olivia Di; Gheorghiu, Vlad; Mosca, Michele: Fault-Tolerant Resource Estimation of Quantum Random-Access Memories. IEEE Transactions on Quantum Engineering, 1:1–13, 2020.

[MSN17] Maier, Alexander; Schriegel, Sebastian; Niggemann, Oliver: Big Data and Machine Learning for the Smart Factory—Solutions for Condition Monitoring, Diagnosis and Optimization. In: Industrial Internet of Things, Springer Series in Wireless Technology, pp. 473–485. Springer International Publishing, 2017.

[NC11] Nielsen, Michael A.; Chuang, Isaac L.: Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press, 10th edition, 2011.

[QCH18] Qader, Mohiuddin Abdul; Cheng, Shiwen; Hristidis, Vagelis: A Comparative Study of Secondary Indexing Techniques in LSM-Based NoSQL Databases. In: Proceedings of the 2018 International Conference on Management of Data. SIGMOD '18. Association for Computing Machinery, p. 551–566, 2018.

[SW19] Sawerwain, Marek; Wróblewski, Marek: Application of Quantum k-NN and Grover's Algorithms for Recommendation Big-Data System. In: Information Systems Architecture and Technology: Proceedings of 39th International Conference on Information Systems Architecture and Technology – ISAT 2018. Springer International Publishing, pp. 235–244, 2019.

[Wa18] Wang, Junping; Zhang, Wensheng; Shi, Youkang; Duan, Shihui; Liu, Jin: Industrial Big Data Analytics: Challenges, Methodologies, and Applications. CoRR, abs/1807.01016, 2018.

[We20] Weigold, Manuela; Barzen, Johanna; Leymann, Frank; Salm, Marie: Data Encodinbg Patterns for Quantum Computing. In: Proceedings of the 27th Conference on Pattern Languages of Programs. PLoP '20. The Hillside Group, 2020.

[We21] Weigold, Manuela; Barzen, Johanna; Leymann, Frank; Salm, Marie: Expanding Data Encoding Patterns For Quantum Algorithms. In: 2021 IEEE 18th International Conference on Software Architecture Companion (ICSA-C). IEEE, pp. 95–101, 2021.

[ZS22] Zajac, Markus; Störl, Uta: Towards quantum-based Search for industrial Data-driven Services. In: Proceedings of the 2022 IEEE International Conference on Quantum Software (QSW). IEEE, 2022.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1101

Programming language for unified computing with classical and quantum bits

Gergely Gálfi[1], Tamás Kozsik[2], Zoltán Zimborás[3]

Abstract: As the capabilities of quantum computers - regarding the numbers of physical qubits, the number of consequtive coherence-keeping steps, or the noise of the gates - are continuously improving, so they enable the implementation of complex algorithms, including ones which combine classical and quantum-enabled steps. Such computations will need control and data shared between the two worlds (quantum and classical), and for improving code quality by eliminating redundancies they may also need a form of classical/quantum polymorphism. This paper proposes a high-level programming language, Qubla to support this type of polymorphism, allowing not only genericity, but also mixed classical and quantum arithmetics. In this language, arithmetical and logical operations can be performed on data formed of both classical bits and qubits. The current implementation of the language works as a interpreter/compiler hybrid: all the classical steps of the program are executed consecutively and together with them the quantum steps are added as building blocks to a quantum logic definition over qubits.

1 Introduction

Many of the quantum algorithms have steps which are quantized version of some common classical procedures. Typical example is Shor's algorithm[Sh94], where to prepare the incoming state of the QFT, controlled modular multiplications of a quantum and classical integers are employed. These steps are usually considered as trivial, in the sense that they require "only" the reimplementation of classical logical circuits with quantum gates. On the other hand, with the increasing number of logical qubits this originally straightforward task of quantum circuit design is becoming far too complex to effectively done manually. This was our main motivation for developing a programming language which allows users to specify their algorithms through constructions similar to the ones in the classical procedural programming languages. Also we wanted to make it transparent as much as possible whether in a given algorithm, classical or quantum bits are used. Certainly there are other programming languages which provide this type of transparency, e.g. Silq[Bi19]. However, we had some specific needs which made us to develop a compiler from scratch instead of using or tweaking an existing one. It is best to compare our language with Silq to shed light on our drivers during the development process:

[1] ELTE Eötvös Loránd University, Budapest, Hungary, galfi@inf.elte.hu, **0000-0002-6568-877X**
[2] ELTE Eötvös Loránd University, Budapest, Hungary, kto@inf.elte.hu, **0000-0003-4484-9172**
[3] Wigner Research Centre for Physics, Budapest, Hungary , zimboras.zoltan@wigner.hu, **0000-0002-2184-526X**

doi:10.18420/inf2022_93

- There are simulation methods or other numerical analysis tasks around quantum computing which require to pass multiple times through the chain of unitary transformations making up the quantum logic itself. In these cases it is more effective to calculate all the transformations in advance. That's why we needed a system which does not simulate or directly run the quantum steps on a physical hardvare during compilation time - as Silq does -, but rather builds up the quantum logic as a whole. It was our definite intention to separate the task of building up the logical structure of a quantum circuit based on a specified algorithm and the task of running that quantum circuit either on a real hardware or on a simulator. As our framework focuses on the former task, the main final product of our compiler is a quantum logic definition, so it was also practical to create it as a library and not a command-line tool. Further on we will refer to this framework as Qubla[22]. Our choice of environment for implementing it was Python, as it gives many opportunity to feed the resulting quantum logics into the very extensive QC ecosystem in Python.
- Also we wanted to focus our attention on transformations which are boolean or - from classical logical engineering standpoint - could be described by truth tables. There are theorems which ensure that any boolean function could be built up from some universal gates (e.g. NAND-gates), but even with a more practical or ergonomical approach, it is common to define these functions as a sequence of some custom few-bit boolean operations. So one of our goal was to get even the most basic operations as addition, multiplication, division, etc. defined by a sequence of truth table operations, and making these operations partially or entirelly quantized by substituting qubits in place of classical bits appropriately. Turning a classical boolean transformation into a quantum one requires to fulfill some mathematical constraints (like the unitarity of the operation), but we've seen this "housekeeping" as a task should be done automatically by the compiler and wanted to create the framework as transparent as possible for this classical-quantum transformation.
- Having the truth table transformations as a separate entity group - and not seeing them as just some general unitary transformations - helps us on one hand to produce a more optimized quantum logic (regarding the number of qubits and steps). On the other hand it could also help the physical implementation of these transformations: Solovay-Kitaev theorem[Ki97] only ensures the existence of an approximate implementation for general unitary transformations however keeping in eyesight that these are boolean ones it is guaranteed that they could be generated exactly by a finite number of elementary operations.

Summarily, we wanted to develop a framework which can process the source code in a hybrid manner: should the users operate with classical quantities (bits, fixed length words or integers), then the instructions are executed on the spot. However, when a user-defined function encounters a qubit-based quantity, it adds a new steps to the quantum logic definition. In the typical case of mixed-type arguments, it is up to the compiler to do in advance as many precomputations as possible, and add only the minium necessary qubit-operations to

the quantum logic definition. So, the final result after processing the source code is the usual output of the classical steps (e.g. lines printed on the console), but additionally a quantum logic definition is generated.

2 Object types

To understand how Qubla works as a unified classical-quantum language, it is worth taking a look at first some of the available data types. Qubla is a dynamically typed language. Converting object to a given type is easy, could be done with applying the desired type as a single-argument function. For example, `int(x)` will convert object `x` to an `int`, provided by the original type of `int(x)` allows that.

2.1 Integer type

Similar to other programming languages, Qubla uses an integer type. Our intention was to avoid any bit number constraint in this language, so the integer type is arbitray-length. To avoid the quirks of indefinite length quantum objects, integer type don't have quantum version, for that purpose the fixed length word type should be used.

2.2 Bit and qubit type

The most important atomic data types of the language are the bit and qubit types. The bits are the usual binary values, the two distinct values could be accessed by `bit(0)` and `bit(1)`. The quantum counterpart of bit objects are the `qbit` objects. These are representing the qubits of the quantum computer. Technically a `qbit` object contains only a reference to a certain qubit, e.g. `qbit[5]` is a reference to the fifth qubit of the quantum computer. As it is clear from this definition, the compiler - apart from the initialization - doesn't keep track of the time evolution of qubits as we consider it a task for a simulator or for a real hardware implementation. However, the compiler guarantees that when referenced multiple times, the qubits wouldn't change their values. This may sounds too restrictive, however this condition is the quantum correspondent of the classical principle that for example `bit(1)` ought to have the same meaning while the program is running. We say that a qubit is *invariable* under a series of quantum steps, if in the case these steps applied on any pure canonical base state, then the density matrix of the qubit doesn't change.

2.3 Word types

To handle more complex (semi)quantum integer-like numbers, Qubla provides fixed length word types, specifically `word` and `uword`. Both of them refers to a finite list of bits and

qubits, and the only difference between the two is how certain functions - e.g. the “less than” function - handle them: `uword` is an unsigned integer, `word` should be considered as a two’s complement signed value. Managing of word objects are relatively easy in Qubla, even when it is a mix of classical bits and qubits. For example, `uword{[bit,bit,qbit,qbit]}(13)` will result in a word object similar to `uword{[1,0,qbit[4],qbit[5]]}` (actual indexing of qubits depends on the preceding part of the source code and also on the state of the compiler), where `qbit[4]` and `qbit[5]` are newly allocated (not used before) qubits and both of them are initialized to the state $|1\rangle$.

3 Quantum logic steps

The following building blocks are employed by Qubla:

- **Initialization steps** As the name suggests, this type of step is responsible for setting one or more qubits to a given quantum state.
- **Truth table steps** These are special permutation transformations where the mapping of the canonical basis states could be defined through the usual truth tables. Certainly not every truth table leads to a permutation, and henceforth a unitary transformation (e.g. the two bit AND operation). However, it is guaranteed by the Qubla compiler, that for any truth table, with arbitrary inputs and outputs, it will be extended by the inclusion of input qubits into the output till unitarity is achieved required by any reversible computing model[To80].
- **Copy steps** Qubla also guarantees the invariability of qubits in the sense of 2.2. To achieve that goal it duplicates the input qubits certainly only in the classical sense, as could be done by a CNOT gate with the controlled input set to zero. So copy steps would well fit into the group of general truth table transformations, however during reduction phase some of these copy steps are to be removed, and it is easier to find them if they are already belong to their own group.

4 Test case

Despite this use case doesn’t have any practical usage in Quantum Computing (it’s output could be simply found out), it was able to drive the Qubla framework through all it’s compilation and reduction steps in a non-trivial way, so giving it an end-to-end testing. Additionally it is a good example on how easily a complex quantum logic could be defined in Qubla. In this text case we define a quantum logic to calculate all the possible products of two 3-bit unsigned word (practically the numbers 0-7). The actual Qubla-code which defines this logic is the following:

```
H = {0: 1/sqrt(2), 1: 1/sqrt(2)};
```

```
n = 3;
arrqbx = alloc(n); arrqby = alloc(n);
for(i : seq(n)){
    arrqbx[i] = qstate(H)[0];
    arrqby[i] = qstate(H)[0];
}
x = uword{n}(arrqbx); y = uword{n}(arrqby);
z = x*y;
output(z);
```

For variables `x` and `y` we create 2x3 qubits, all being in equal superposition (so all possible input numbers has equal probability). The statement `z = x*y` calculates the 6-bit arithmetic product of these inputs. The ultimate line `output(z)` makes it sure that the qubits corresponding to `z` and all other steps leading up to it will not be removed in the reduction phase. After compiling and reducing the quantum logic, we ended up in a 21-qubit system. This is well within the capability of a simple state vector simulation. So after simulating this system and then calculating the probabilities of the possible values of `z`, we were able to compare it with the theoretically predictable distribution (the distribution of the product of two numbers between 0 and 7, picked in an uniformly random way). Having an exact match give us a strong evidence that each individual steps in our Qubla framework work correctly.

4.0.1 Acknowledgements

This project has received funding from the Ministry of Innovation and Technology and the National Research, Development and Innovation Office within the Quantum Information National Laboratory of Hungary and through grant No. FK135220.

References

[22] Qubla source code, 2022, URL: http://absimp.org/qubla.

[Bi19] Bichsel, B.; Baader, M.; Gehr, T.; Vechev, M.: Silq: A High-level Quantum Programming Language, Sept. 2019, URL: http://silq.ethz.ch/, visited on: 09/22/2019.

[Ki97] Kitaev, A. Y.: Quantum computations: algorithms and error correction. Russian Mathematical Surveys 52/6, pp. 1191–1249, Dec. 1997, URL: https://doi.org/10.1070/rm1997v052n06abeh002155.

[Sh94] Shor, P. W.: Algorithms for Quantum Computation: Discrete Logarithms and Factoring, 1994.

[To80] Toffoli, T.: Reversible computing. International Colloquium on Automata, Languages, and Programming/, pp. 632–644, 1980.

GI Quantum Computing Workshop

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1109

Mapping Quantum Circuits to 2-Dimensional Quantum Architectures

Kamalika Datta[1], Abhoy Kole[2], Indranil Sengupta[3], Rolf Drechsler[4]

Abstract: We have been witnessing a rapid growth in quantum computing research over the years, with the emergence of demonstrable quantum computers of moderate size. The major issues that are faced to run a quantum algorithm reliably on these systems are: (i) lower qubit coherence period, (ii) noisy primitive gate operations, (iii) limited number of available physical qubits, and (iv) support of restricted set of 2-qubit operations. Overcoming these issues mandates physical resources that exceeds the capabilities of these noisy intermediate scale quantum (NISQ) systems. On the other hand, computation using bare qubits get further disturbed due to the inclusion of additional gates to mitigate the nearest neighbor constraints. In the present work, the 2-dimensional square, heavy-hex and fully hexagonal qubit coupling lattices are considered for mapping quantum circuits. The benefits are assessed in terms of minimal additional gates needed to satisfy the nearest neighbor (NN) constraint and the compilation complexity of mapping circuits on these architectures. From the experiments by mapping benchmark circuits on 16-qubit square and 65-qubit heavy-hex architectures from IBM as well as on a 64-qubit fully hexagonal architecture, it is observed that none of the square or heavy-hex lattice architecture provides uniform compilation advantage compared to the fully hexagonal architecture. It is expected that beyond the NISQ era, strongly connected lattices like hexagonal will become practically feasible.

1 Introduction

With the emergence of quantum computers in recent years, researchers have started to work on application mapping with respect to more realistic qubit architectures [IB]. The size of quantum computers in terms of the number of qubits has been steadily increasing over the years. For example, IBM has recently announced their 127-qubit Eagle processor [CDG21]. In quantum computing, qubits are used as the fundamental unit of information as opposed to bits in classical computing. The qubits can be regarded as hardware, on which the quantum gates operate in the axis of time.

[1] German Research Centre for Artificial Intelligence (DFKI), Cyber Physical System, Bremen, Germany
kamalika.datta@dfki.de
[2] German Research Centre for Artificial Intelligence (DFKI), Cyber Physical System, Bremen, Germany
abhoy.kole@dfki.de
[3] JIS University, Department of Computer Science and Engineering, Kolkata, India,
Indian Institute of Technology Kharagpur, Department of Computer Science and Engineering, India
indranil.sengupta@jisuniversity.ac.in
[4] University of Bremen, Group of Computer Architecture, Bremen, Germany,
German Research Centre for Artificial Intelligence (DFKI), Cyber Physical System, Bremen, Germany
drechsler@uni-bremen.de

doi:10.18420/inf2022_94

In a quantum computer, the qubit architecture defines how the qubits are laid out physically and the ways they can interact. The concept of *coupling map* has been used in the earlier generations of IBM Q series of quantum computers [IB], which specifies the way the physical qubits can interact during 2-qubit gate operations. Some other qubit architectures have also been proposed in recent years, like the *hexagonal architecture* [Li17, Ta18], where a qubit can be connected with up to six other neighboring qubits in a two dimensional (2-D) grid. In fact, all practical qubit architectures consider the layout of the qubits on a 2-D plane with realistic interconnections. The number of neighborhood connections to a qubit is determined by factors like physical layout, errors during gate operations, etc. [Na21]. One of the basic constraint in mapping a quantum circuit to these physical architectures is that the interacting qubits must be neighbors. For a large quantum circuit with fewer physical connectivity amongst qubits, it is indeed a major challenge to satisfy this constraint. In most of the cases, *Swap* gates are inserted to exchange the states of two qubits, where each Swap gate is implemented using three back-to-back CNOT gates. The consequent increase in the number of gates in the circuit makes it more prone to errors, and hence reducing the number of Swap gate operations is of vital importance.

In the present NISQ era, with the availability of limited number of noisy qubits, the degree of association of the qubits is limited to 2 or 3 (e.g., 2-D and heavy-hex architectures). However, with further developments in fabrication technology, emergence of strongly connected qubit architectures (e.g., hexagonal) is a strong possibility. The main objective of the present work is to compare some of the recent qubit architectures with the hexagonal architecture with respect to the cost of mapping quantum circuits, and evaluate how much reduction in gate overheads can be achieved.

In this paper we analyze the performances of 16-qubit IBM QX5 (Rueschlikon) and 65-qubit IBM Q65 (Hummingbird) architectures, as well as the hexagonal architecture. We propose methods to map a given quantum circuit to these architectures. The rest of the paper is organized as follows. Section 2 provides a survey of the selected IBM and the hexagonal architectures. Section 3 proposes the qubit mapping algorithms for the architectures selected for comparison. In section 4 we provide the experimental results followed by concluding remarks in section 5.

2 Recent Quantum Architectures

In this section we provide a brief review of the IBM QX5 and Q65 architectures and also the 2-D hexagonal architecture. The IBM Q series of quantum computers are based on qubits that are built using superconducting transmon, using materials like niobium and aluminum, patterned on a silicon substrate. IBM first came up with 5-qubit chips in 2016 and made it available in the cloud. Since then, IBM came up with several quantum chips of increasing capacity, like 16-qubit Canary and Rueschlikon, 27-qubit Falcon, 65-qubit Hummingbird, and 127-qubit Eagle [IB17]. In this paper we particularly focus on IBM's QX5 (Rueschlikon) and Q65 (Hummingbird) architectures. Earlier qubit architectures by IBM like the QX5

relied on *2-D lattice* structure. In contrast, the Q65 architecture uses the *heavy-hex lattice* structure, which is the topology used in recent IBM quantum computers. The choice of such topology is based on experimental evaluation to ensure reduced error-rates, which also provides room to explore error correcting codes. In these topologies, a single qubit is connected with a maximum of two or three neighboring qubits.

The IBM Q series architectures support a limited number of 2-qubit gate operations. Fig. 1 shows the qubit arrangements in the QX5 [Ga] and Q65 [IB20] architectures. An edge denotes coupling between physical qubits for conducting 2-qubit gate operations (e.g. CNOT) on such architectures. The QX5 architecture (see Fig. 1a) is more restrictive, where the edge directions indicate control and target qubits for 2-qubit CNOT operations, e.g. for a directed edge like $q_i \rightarrow q_j$ between physical qubits q_i and q_j, the QX5 architecture supports CNOT(q_i, q_j) operation, whereas it needs four additional Hadamard operations to conduct CNOT(q_j, q_i) operation [ZPW19].

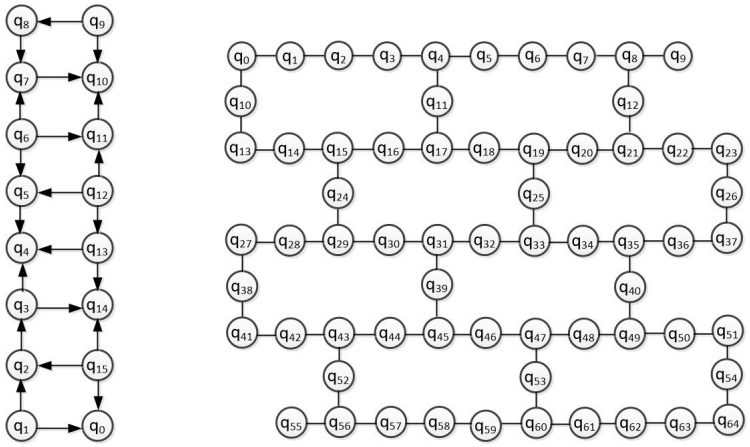


Fig. 1: (a) IBM QX5 Rueschlikon architecture, (b) IBM Q65 Hummingbird architecture.

Recently quantum mechanical systems based on hexagonal architecture have been proposed [Li17, Ta18]. In this architecture, each qubit can have a maximum of 6 neighbors. Having more connectivity amongst neighbors provides flexibility in the mapping of quantum circuits. Fig. 2 shows a 6×5 hexagonal architecture where the height is 5 and width is 6. Recently few works have been reported on mapping quantum circuits to hexagonal structures [CL21, Da22]. As mentioned earlier, Swap gates are typically inserted to make a quantum circuit NN-compliant. In [CL21], Chang et al. have used Swap gates to make quantum circuits NN-compliant. Due to increased connectivity among qubits, the hexagonal architecture is found to reduce the number of Swap gates required as compared to existing architectures. In [Da22], Datta et al. have used the CNOT templates [RD12] as the cost metric for NN-compliance in hexagonal architecture, instead of Swap gates. The method

uses a Genetic Algorithm based global ordering of qubits for generating the initial qubit placement, and is found to outperform other NN-mapping techniques in two dimensions.

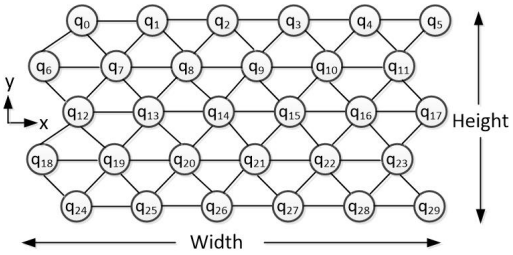


Fig. 2: Hexagonal qubit architecture.

It is evident that IBM's QX5 and Q65 architectures have less connectivity as compared to hexagonal architecture, and therefore have distinct advantages in terms of overall error rate. Although hexagonal architecture seems promising with higher connectivity, commercial products based on such architecture are expected only in the post-NISQ era.

3 Mapping Quantum Circuits to Various Architectures

In this section we discuss how the logical qubits in a quantum circuit can be mapped to various qubit architectures. For the sake of comparison, we have used Swap gates as the evaluation metric for all the methods. The overall flow of quantum circuit mapping is shown in Fig. 3. We first discuss the steps for qubit mapping in the Q65 architecture; the mapping for the QX5 architecture is very similar.

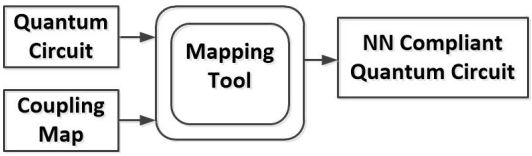


Fig. 3: The overall mapping flow.

3.1 Mapping Quantum Circuits to IBM's 65-bit Hummingbird Architecture

The problem of mapping quantum circuits to IBMQ architectures has been explored by several researchers [ZPW19, dADdS19, Ko20, Ni20, NBD21]. In most of these methods either 5-qubit (e.g., QX2 (Yorktown) and QX4 (Tenerife)) or 16-qubit (e.g., QX3 and QX5) Rueschlikon architectures have been considered. In this paper we have used QX5 as well as Q65 architectures for experimentation. A generalized architecture-independent qubit mapping tool has been developed for the purpose. The inputs to the tool are the

architectural specification of the qubits (e.g., Q65), and the given quantum circuit. The algorithm generates the qubit placement, and calculates the number of Swap gates required to make the circuit NN-compliant. When we map a circuit with n logical qubits $(Q_0, \ldots, Q_{n-1})$ to the Q65 architecture, the n physical qubits can be selected in $\binom{65}{n}$ ways. It may be noted that the qubit association is bidirectional in Q65, i.e. an edge (q_i, q_j) in the coupling map indicates that either q_i or q_j can be used as target in a 2-qubit gate operation. Also, the degree of association of any qubit is 2 or 3. Referring to Fig. 1b, the qubit q_4 has the connections $q_4 \rightarrow q_3$, $q_4 \rightarrow q_5$, and $q_4 \rightarrow q_{11}$, and hence the degree of association is $deg(q_4) = 3$. Considering a quantum circuit, the degree of association is represented by connections between the logical qubits in the qubit interaction graph, $QIB = (V, E)$, where the vertices $Q_k \in V$ indicate logical qubits, and the weight (w_{kr}) of an edge $(Q_k, Q_r) \in E$ denotes the number of 2-qubit gate operations between Q_k and Q_r.

The following steps are used to map a quantum circuit to the IBM Q65 architecture. The process of mapping to the QX5 architecture is similar, with special care taken for the directional constraints in the coupling map.

(i) For every pair of physical qubits (q_i, q_j), d_{ij} denotes their distance as per the coupling map. If the qubits are adjacent, we set $d_{ij} = 0$. If they are non-adjacent, the distance is initially set to infinity (∞). To find the shortest distance between all pairs of non-adjacent qubits, we use Floyd-Warshall algorithm, which has an overall runtime of $O(n^3)$ for n physical qubits. This process needs to be performed only once for an architecture and the computed distance measure (d_{ij}) can be kept in a persistent storage to speed up the mapping process.

(ii) A greedy approach is used to select the n physical qubits to map a n-qubit quantum circuit. The objective is two-fold: (i) to minimize the number of CNOT operations on uncoupled physical qubits in the mapped quantum circuit, and (ii) to reduce the search time for selecting a good set of n among m physical qubits for mapping. In the first step of qubit mapping, the logical qubit (Q_k) having highest degree of association is mapped to a physical qubit (q_i) that also has highest degree of association. In case of a tie, one of the qubits involved in the tie is chosen randomly. This process is repeated for all the remaining logical qubits, in descending orders of degree of association. At every step, the next physical qubit is selected based on the shortest distance from the already assigned qubits. For mapping n physical qubits, the average running time of this step is $O(n^2)$.

(iii) During the mapping of a quantum circuit to a physical qubit architecture, the selection of physical qubits and also the initial permutation of the logical qubits $(Q_0, \ldots, Q_{n-1})$ play an important roles in reducing the number of Swap gates and also the circuit depth. In general, finding the optimal initial permutation for n logical qubits is computationally hard, with a complexity of $O(n!)$. In this work we have used an evolutionary algorithm to find the initial permutation of logical qubits, some of the parameters for which are as follows.

- The number of Swap gates required for NN-compliance is used as the *fitness function* consistent with the edge weights of the corresponding QIG.
- The initial population size is taken as 30, crossover rate 70%, and mutation rate 17%. The algorithm is run over 500 generations, at the end of which the best solution gives the initial permutation.
- While generating the chromosomes for the next generation, the best 4 permutations from the current generation are copied. For crossover operation, a pair of permutations (P_i^c, P_j^c) from the current generation are picked out using roulette wheel selection method [Go89]. Using a randomly selected crossover point, first parts of this selected pair are copied to a new pair (P_1^n, P_2^n) for the next generation. To prevent redundant qubit mapping, the remaining missing qubits of the new pair (P_i^n, P_j^n) are filled up by reordering their appearance in the pair (P_i^c, P_j^c). During mutation, one of the operations like swapping positions of two randomly selected qubits, moving a randomly selected qubit to a vacant adjacent position or moving a qubit to a randomly selected vacant position is conducted randomly on permutations from the current generation before being copied in next generation.

(iv) After the initial placement of the logical qubits, some of the 2-qubit gate operations may not satisfy the coupling constraints, for which Swap gates are inserted. A Swap gate exchanges the states of a pair of qubits, which in turn changes the qubit permutation. Starting with the initial mapping, we traverse the quantum circuit from left to right, and use a k-gate lookahead approach to insert Swap gates before each 2-qubit gate operating on non-coupled qubits. The objective is to make the gate operation NN-compliant. Among the various possibilities that may exist, we select the one that minimizes the cost of mapping the k subsequent gates in the circuit.

For a given quantum circuit, the mapping tool tries to carry out a good logical to physical qubit mapping, and report the overhead for NN-compliance in terms of the number of Swap gates required.

3.2 Mapping Quantum Circuits to Hexagonal Architecture

We now discuss the problem of mapping a quantum circuit to the hexagonal qubit architecture. Fig. 2 depicts the coupling map of 30 qubits arranged in a 6×5 hexagonal array, with a maximum qubit coupling of 6. As a matter of convention, we number the qubits in row-major order, from left to right and then from top to bottom. If W and H denote the width and height respectively in the $W \times H$ arrangement of qubits, the total number of qubits will be $W.H$. The Cartesian co-ordinates of the i^{th} qubit (i.e., q_i) can be calculated as

$$\begin{aligned} x_i &= (1 - (y_i \,\%\, 2)) + 2\,(i \,\%\, W) \\ y_i &= i \div W \end{aligned}$$

where % and ÷ respectively denote modulus and integer division. This coordinate system guarantees that the Manhattan Distance (MD) between any two neighboring qubits is 2. The MD measure can be used to obtain the distance d_{ij} between physical qubits q_i and q_j as:

$$d_{ij} = max\left(|y_i - y_j|, \frac{MD(q_i, q_j)}{2}\right) - 1 \quad (1)$$

where the coordinates of the qubits q_i and q_j are (x_i, y_i) and (x_j, y_j) respectively.

The mapping approach is similar to the previously outlined approach for IBMQ architecture with minor changes. The following steps are followed to map the logical qubits of a quantum circuit to the hexagonal architecture:

(i) We create a weighted qubit interaction graph (QIG) from the given quantum circuit, in which every node Q_k denotes a logical qubit and an edge (Q_k, Q_r) denotes the presence of 2-qubit gate(s) between Q_k and Q_r. The weight of the edge w_{kr} indicates the number of 2-qubit gates between the pair of qubits.

(ii) We use an evolutionary algorithm to find the best initial mapping of the qubits. The earlier defined cost function (viz., the total number of Swap operation for obtaining NN-compliance) is used for optimization keeping population size, crossover rate, mutation rate and number of generations unchanged.

(iii) With the obtained initial mapping, the quantum circuit from left to right is traversed using the same k-gate lookahead approach, and Swap gates are inserted before each 2-qubit gate operating on non-coupled qubits for NN-compliance. Different possibilities are explored and the lowest cost alternative is selected for mapping the k subsequent gates in the circuit.

4 Experimental Results

In this section we present the results of mapping various quantum benchmark circuits to the selected IBM and hexagonal qubit architectures using the approaches outlined in the previous section. Initially, the RevLib benchmarks [Wi08] available in reversible form are transpiled using Qiskit 0.19.0 [An21] with Python 3.16.12 setting the Qiskit parameters *basis_gates* and *optimiyation_levels* as (CX, U3) and 0, respectively before being used in our experiments. The qubit mapping algorithms are implemented using C++ and run on an Intel(R) Xeon(R) Gold 6132 processor with 96GB memory running CentOS 7.0 operating system.

We have shown two sets of experimental results. In Table 1, we show the mapping results for IBM QX5 and IBM Q65. The first three columns show the benchmark names, number of qubits (n), and number of CNOT gates in the original netlist. The next three columns

show the number of extra CNOT gates required to make the circuit NN-compliant, the number of H-gates (specifically required to conduct CNOT operation consistent with the coupling direction), circuit depth and run-time for IBM QX5. The next three columns show the number of extra CNOT gates required, circuit depth and run-time for IBM Q65. The last three columns show the better architecture, % improvements in number of gates, and difference in run-times. We have chosen 51 benchmarks with $3 \leq n \leq 14$, out of which 21 benchmarks show better results for QX5, and 30 show better results for Q65. Out of the total 21 benchmarks for which QX5 gives better results, 11 of them have $n \geq 10$. But out of 30 benchmarks for which Q65 provides better results, most of them have $3 \leq n \leq 7$. It may be noted that in the QX5 architecture, the degree of association is 2 for only 4 qubits, and 3 for the remaining 12 qubits. In contrast for Q65, considering the first 16 qubits, 3 of them have degree of association as 3, while the rest have a value of 2. However, the Q65 architecture has better error resiliency and can map larger quantum circuits as compared to QX5.

Table 2 shows the mapping results for 8×8 hexagonal and IBM Q65 architectures. The first three columns show the benchmark names, number of qubits (n) and number of CNOT gates. The next six columns show the number of extra CNOT gates required for NN-compliance, circuit depth and run-time for the hexagonal and Q65 architectures respectively. The last two columns show the % improvement in CNOT gates and difference in run-time of the hexagonal architecture over Q65. It is evident from the results that the hexagonal architecture provides lower mapping overheads as compared to Q65, which is a result of the extended qubit coupling in the hexagonal architecture. However, it may be noted that the hexagonal architecture is not practically feasible in the present NISQ era; however, it can be a potential candidate in the future during the post-NISQ era.

5 Conclusion

In this paper we have analyzed the overheads of quantum circuit mapping on two IBM's quantum computer architectures, viz. 16-qubit Rueschlikon (QX5) and 65-qubit Hummingbird (QX65), and a 2-D hexagonal architecture. Both QX5 and Q65 architectures have a maximum qubit neighborhood of 3, with directional constraints existing between pair of coupled qubits in QX5 and no directional constraints in Q65. In fact, in the QX5 the qubits are laid out uniformly in a 2-D grid, while a heavy-hex lattice structure is used to connect the qubits in Q65. In contrast, the hexagonal architecture has a maximum qubit neighborhood of 6. Experimental results reveal that the hexagonal architecture incurs less gate overhead for mapping various benchmarks compared to QX5 and Q65, which is mainly because of the extended connectivity among qubits. Between QX5 and Q65, since the qubit interaction patterns are different, we see mixed results for the mapping of benchmark circuits. It may be noted that although hexagonal architecture provides better mapping, cloud based devices based on such extended neighborhood architectures are yet not available. In comparison, the IBM series of quantum computers in spite of their limitations have been physically fabricated and widely used by users from all over the globe. Hexagonal and other strongly connected architectures may emerge in the future during the post-NISQ era.

Tab. 1: Comparison of NN-mapping on IBM QX5 and Q65 architectures.

Benchmark Circuits			QX5				Q65			Improvement		
Name	n	#CNOT	#CNOT	#H	#Depth	t(s)	#CNOT	#Depth	t(s)	Arch.	$\Delta_g(\%)$	$\Delta_t(s)$
3_17_14	3	15	15	24	57	0.05	9	39	0.27	Q65	40.00	−0.22
4_49_17	4	35	21	72	122	0.08	27	92	0.42	QX5	−28.57	−0.34
4gt10-v1_81	5	49	39	156	167	0.09	36	113	0.55	Q65	7.69	−0.46
4gt12-v0_86	5	81	66	228	282	0.10	54	167	0.52	Q65	18.18	−0.42
4gt13_90	5	43	33	112	156	0.11	30	98	0.57	Q65	9.09	−0.46
4gt13_91	5	39	30	104	145	0.10	27	91	0.56	Q65	10.00	−0.46
4gt4-v0_73	5	127	102	348	408	0.11	90	269	0.54	Q65	11.76	−0.44
4gt5_77	5	42	30	128	135	0.10	12	84	0.54	Q65	60.00	−0.44
4mod5-bdd_287	7	27	27	52	89	0.15	30	68	0.78	QX5	−11.11	−0.64
4mod5-v1_23	5	28	27	60	93	0.10	27	73	0.54	Q65	0.00	−0.45
4mod7-v0_94	5	54	27	76	154	0.10	30	126	0.56	QX5	−11.11	−0.46
aj-e11_165	4	56	36	144	175	0.07	39	139	0.38	QX5	−8.33	−0.30
aj-e11_168	4	38	24	80	102	0.08	24	86	0.42	Q65	0.00	−0.34
alu-bdd_288	7	33	30	52	94	0.14	24	74	0.71	Q65	20.00	−0.58
alu-v2_30	5	160	126	432	528	0.10	120	363	0.55	Q65	4.76	−0.45
alu-v2_31	5	146	108	364	486	0.12	96	326	0.57	Q65	11.11	−0.45
alu-v4_36	5	40	42	92	131	0.10	39	112	0.5	Q65	7.14	−0.40
C17_204	7	156	132	292	463	0.14	189	355	0.71	QX5	−43.18	−0.58
cm42a_207	14	593	588	1420	1890	0.32	660	1286	1.27	QX5	−12.24	−0.96
dc1_221	11	642	633	1520	2110	0.30	696	1485	1.19	QX5	−9.95	−0.89
decod24-enable_126	6	116	111	328	430	0.13	108	283	0.66	Q65	2.70	−0.53
decod24-v1_41	4	29	24	72	91	0.08	12	60	0.41	Q65	50.00	−0.33
decod24-v3_45	4	47	36	132	154	0.08	18	103	0.4	Q65	50.00	−0.33
ex3_229	6	114	102	340	412	0.13	90	251	0.67	Q65	11.76	−0.54
ham7_104	7	114	108	296	387	0.13	102	261	0.64	Q65	5.56	−0.51
hwb4_49	4	85	54	204	260	0.07	45	197	0.36	Q65	16.67	−0.29
hwb5_55	5	141	141	416	505	0.11	126	325	0.6	Q65	10.64	−0.48
hwb6_58	6	183	201	512	670	0.13	198	452	0.66	Q65	1.49	−0.53
hwb7_61	7	6252	5718	14 516	19 706	4.43	5628	13 518	1.43	Q65	1.57	3.00
hwb7_62	7	4142	3828	9840	12 847	2.07	3864	8736	1.17	QX5	−0.94	0.90
mini_alu_305	10	67	84	148	180	0.20	63	134	1.03	Q65	25.00	−0.83
mod5adder_127	6	165	165	476	562	0.12	150	398	0.61	Q65	9.09	−0.49
mod5adder_128	6	110	87	276	404	0.12	93	261	0.65	QX5	−6.90	−0.52
mod8-10_177	5	133	129	324	482	0.11	96	298	0.57	Q65	25.58	−0.46
mod8-10_178	5	99	78	288	319	0.11	78	235	0.56	Q65	0.00	−0.45
one-two-three-v0_97	5	98	87	276	354	0.10	81	243	0.54	Q65	6.90	−0.43
one-two-three-v1_99	5	47	48	152	173	0.10	45	123	0.52	Q65	6.25	−0.42
pm1_249	14	593	588	1348	1857	0.32	618	1356	1.29	QX5	−5.10	−0.97
rd32_273	5	160	93	372	472	0.11	138	361	0.5	QX5	−48.39	−0.39
rd53_131	7	160	135	252	422	0.12	141	361	0.63	QX5	−4.44	−0.50
rd53_135	7	108	99	216	329	0.14	114	256	0.75	QX5	−15.15	−0.60
rd73_140	10	90	90	212	270	0.20	87	193	1.02	Q65	3.33	−0.82
rd73_252	10	2380	2094	4588	6635	0.89	2730	5041	1.4	QX5	−30.37	−0.51
sqn_258	10	4726	4191	8760	12 173	2.32	5172	10 672	1.66	QX5	−23.41	0.67
sqrt8_260	12	1711	1512	3032	4444	0.58	1785	3676	1.42	QX5	−18.06	−0.84
sym6_145	7	1302	1188	3056	4037	0.36	1023	2708	0.75	Q65	13.89	−0.38
sym6_316	14	107	138	280	326	0.24	210	300	1.2	QX5	−52.17	−0.96
sym9_146	12	128	69	308	306	0.23	138	257	1.22	QX5	−100.00	−0.99
sym9_148	10	7224	6273	16 560	22 996	6.40	6711	15 670	2	QX5	−6.98	4.41
wim_266	11	326	324	828	1067	0.21	354	788	0.94	QX5	−9.26	−0.73
z4ml_269	11	1248	1131	2636	3589	0.43	1533	2827	1.2	QX5	−35.54	−0.77

Tab. 2: Comparison of NN-mapping on hexagonal 64-qubit and IBM Q65 architectures.

Benchmark Circuits			Hexagonal (8×8)			IBM Q65			Improvement	
Name	n	#CNOT	#CNOT	#Depth	t(s)	#CNOT	#Depth	t(s)	Δ_g (%)	Δ_t (s)
4_49_16	4	78	12	140	0.34	51	189	0.41	76.47	0.07
4gt12-v0_86	5	81	24	159	0.49	54	167	0.52	55.56	0.03
4mod5-v0_20	5	9	0	13	0.00	6	19	0.44	100.00	0.44
4mod5-v1_22	5	10	0	13	0.00	6	19	0.45	100.00	0.45
4mod5-v1_23	5	28	6	53	0.41	27	73	0.54	77.78	0.13
5xp1_194	17	4777	1494	8448	3.09	5649	10 214	2.79	73.55	−0.30
add6_196	19	24 745	7266	44 041	8.32	32 499	54 162	12.21	77.64	3.89
alu-v2_30	5	160	57	309	0.51	120	363	0.53	52.50	0.02
alu-v2_31	5	146	27	270	0.49	96	326	0.56	71.88	0.07
alu1_198	20	324	180	454	2.54	438	541	1.39	58.90	−1.15
alu3_200	18	17 787	4032	30 709	5.32	20 376	37 210	7.46	80.21	2.14
C7552_205	21	3301	1212	6011	5.03	4458	7529	3.69	72.81	−1.34
cm150a_210	22	3653	954	6558	5.59	3735	8071	3.45	74.46	−2.14
cm163a_213	29	4911	1110	8722	7.23	5802	10 427	7.54	80.87	0.31
cu_219	25	16 396	3381	28 280	9.73	21 339	38 623	16.14	84.16	6.41
cycle10_293	39	222	327	463	6.33	693	586	4.10	52.81	−2.23
decod_217	21	3301	1284	6019	4.84	4185	7187	3.66	69.32	−1.18
dk17_224	21	13 154	3096	23 029	7.45	19 125	30 663	10.10	83.81	2.65
dk27_225	17	725	297	1281	2.86	1146	1849	2.03	74.08	−0.83
ham15_298	45	313	573	598	9.09	40 722	40 234	5.56	98.59	−3.53
ham7_299	21	151	126	294	2.93	318	427	1.73	60.38	−1.20
hwb6_301	46	589	804	1026	12.50	1863	1529	6.72	56.84	−5.78
hwb7_59	7	8423	1815	15 466	1.32	7260	18 500	1.79	75.00	0.47
hwb8_116	8	12 245	3843	22 708	1.97	13 101	26 210	2.70	70.67	0.73
inc_237	16	5582	1845	9923	3.19	360 666	366 212	5.12	99.49	1.93
life_238	10	19 936	3831	35 063	2.91	20 799	43 599	4.14	81.58	1.23
max46_240	10	22 924	4230	40 067	3.07	22 815	48 798	4.56	81.46	1.49
mod5adder_306	32	337	333	564	6.64	867	815	3.19	61.59	−3.45
mux_246	22	3652	1074	6585	6.39	3747	7598	3.45	71.34	−2.94
parity_247	17	16	36	28	1.07	87	59	1.39	58.62	0.32
pcler8_248	21	956	345	1597	3.84	1284	2063	2.13	73.13	−1.71
plus127mod8192_308	25	91	78	168	2.72	237	242	1.93	67.09	−0.79
plus63mod4096_309	23	82	66	160	2.53	177	220	1.77	62.71	−0.76
plus63mod8192_310	25	89	75	170	2.19	276	275	1.65	72.83	−0.54
rd73_312	25	246	207	487	4.35	522	651	2.11	60.34	−2.24
rd84_253	12	9677	2769	17 209	3.25	10 338	21 784	3.32	73.22	0.07
rd84_313	34	343	333	504	6.69	777	859	3.59	57.14	−3.10
sqr6_259	17	2292	828	4064	3.46	3114	5086	2.51	73.41	−0.95
sym9_193	10	39 296	7131	68 170	3.92	42 186	81 443	7.46	83.10	3.54
sym9_317	27	240	225	414	4.39	555	593	2.41	59.46	−1.98
t481_263	17	376	141	741	2.24	510	859	1.22	72.35	−1.02
urf1_278	9	24 916	14 307	46 644	5.06	40 344	60 383	7.23	64.54	2.17
urf2_152	8	30 180	18 735	65 365	4.87	39 120	86 416	6.45	52.11	1.58
urf3_279	10	66 435	31 281	121 272	11.01	102 738	160 603	18.21	69.55	7.20
x2_267	17	1990	780	3607	2.96	2439	4632	1.81	68.02	−1.15

Bibliography

[An21] Anis, Md S.; A.-Mitchell; Abraham, H. et al.: , Qiskit: An Open-source Framework for Quantum Computing, 2021. [Accessed: 2021-12-20].

[CDG21] Chow, Jerry; Dial, Oliver; Gambetta, Jay: , IBM Quantum breaks the 100-qubit processor barrier. https://research.ibm.com/blog/127-qubit-quantum-processor-eagle/, 2021. [Online; accessed 16-November-2021].

[CL21] Chang, K. Y.; Lee, C. Y.: Mapping Nearest Neighbor Compliant Quantum Circuits onto a 2-D Hexagonal Architecture. IEEE Trans. on CAD of Integrated Circuits and Systems, pp. 1–14, 2021.

[Da22] Datta, K.; Kole, A.; Sengupta, I.; Drechsler, R.: Nearest Neighbor Mapping of Quantum Circuits to Two-Dimensional Hexagonal Qubit Architecture. In: Int'l Symp. on Multiple-Valued Logic. May 2022.

[dADdS19] de Almeida, A. A. A.; Dueck, G. W.; da Silva, A. C. R.: CNOT gate mappings to Clifford+T circuits in IBM architectures. In: Int'l Symp. on Multiple-Valued Logic. pp. 7–12, 2019.

[Ga] Gambetta, Jay: , IBM QX backend information. https://github.com/Qiskit/ibmq-device-information. [Accessed: 2022-04-20].

[Go89] Goldberg, D.: Genetic algorithms in search, optimization, and machine learning. Addision-Wesley Professional, 1989.

[IB] IBM: , IBM Q. https://www.research.ibm.com/ibm-q/. [Accessed: 2019-03-20].

[IB17] IBM: , IBM Quantum Processor Types. https://quantum-computing.ibm.com/services/docs/services/manage/systems/processors/, 2017.

[IB20] IBM: , The map view of IBM Hummingbird Architecture. https://quantum-computing.ibm.com/services?services=systems&type=Hummingbird&system=ibmq_brooklyn/, 2020.

[Ko20] Kole, A.; Hillmich, S.; Datta, K.; Wille, R.; Sengupta, I.: Improved Mapping of Quantum Circuits to IBM QX Architectures. IEEE Trans. on CAD of Integrated Circuits and Systems, 39(10):2375–2383, 2020.

[Li17] Litinski, D.; Kesselring, M. S.; Eisert, J.; von Oppen, F.: Combining Topological Hardware and Topological Software: Color-Code Quantum Computing with Topological Superconductor Networks. Physical Review, 7(3):031048, 2017.

[Na21] Nation, Paul; Paik, Hanhee; Cross, Andrew; Nazario, Zaira: , The IBM Quantum heavy hex lattice. https://research.ibm.com/blog/heavy-hex-lattice/, 2021. [Online; accessed 7-July-2021].

[NBD21] Niemann, P.; Bandyopadhyay, C.; Drechsler, R.: Combining SWAPs and remote Toffoli gates in the mapping to IBM QX architectures. In: Design Automation and Test in Europe. pp. 1–6, 2021.

[Ni20] Niemann, P.; de Almeida, A. A. A.; Dueck, G.; Drechsler, R.: Design Space Exploration in the Mapping of Reversible Circuits to IBM Quantum Computers. In: 23rd Euromicro Conference on Digital System Design (DSD). pp. 401–407, 2020.

[RD12] Rahman, M.; Dueck, G. W.: Synthesis of Linear Nearest Neighbor Quantum Circuits. In: 10th Int'l Workshop on Boolean Problems. pp. 1–9, 2012.

[Ta18] Tang, H. et al.: Experimental quantum fast hitting on hexagonal graphs. Nature Photonics, 12(12):754–758, 2018.

[Wi08] Wille, R.; Große, D.; Teuber, L.; Dueck, G.W.; Drechsler, R.: RevLib: An Online Resource for Reversible Functions and Reversible Circuits. In: Int'l Symp. on Multi-Valued Logic. pp. 220–225, 2008. RevLib is available at http://www.revlib.org.

[ZPW19] Zulehner, A.; Paler, A.; Wille, R.: An Efficient Methodology for Mapping Quantum Circuits to the IBM QX Architectures. IEEE Trans. on CAD of Integrated Circuits and Systems, 38(7):1226–1236, 2019.

Towards Model-Driven Engineering for Quantum AI

Armin Moin[1] Moharram Challenger[2] Atta Badii[3] Stephan Günnemann[4]

Abstract: Over the past decade, Artificial Intelligence (AI) has provided enormous new possibilities and opportunities, but also new demands and requirements for software systems. In particular, Machine Learning (ML) has proven useful in almost every vertical application domain. In the decade ahead, an unprecedented paradigm shift from classical computing towards Quantum Computing (QC), with perhaps a quantum-classical hybrid model, is expected. We argue that the Model-Driven Engineering (MDE) paradigm can be an enabler and a facilitator, when it comes to the quantum and the quantum-classical hybrid applications. This includes not only automated code generation, but also automated model checking and verification, as well as model analysis in the early design phases, and model-to-model transformations both at the design-time and at the runtime. In this paper, the vision is focused on MDE for Quantum AI, particularly Quantum ML for the Internet of Things (IoT) and smart Cyber-Physical Systems (CPS) applications.

Keywords: model-driven engineering; artificial intelligence; machine learning; quantum computing; cyber-physical systems; internet of things

1 Introduction

In October 2019, Arute et al. [AAea19] published the breakthrough results of their research and development at Google AI Quantum on *Quantum Computing (QC)*, where they reported the total runtime of only 200 seconds for a specific computational task on their programmable superconducting quantum processor, which would take approximately 10,000 years on a state-of-the-art *classical* (i.e., non-quantum) supercomputer. This groundbreaking experiment validated the previously hypothetical and theoretical vision of *quantum supremacy*, initiated by Richard Feynman in the 1980s. Today, it is a realistic expectation that QC will eventually be a revolutionary and disruptive enabler technology in the upcoming years and decades in various domains and disciplines: from cryptography and security to Artificial Intelligence (AI), many pieces of software and technologies will become obsolete. Consequently, the current practitioners, such as software developers and data scientists would - in principle - need to learn the new programming paradigms, APIs, and frameworks for QC, in order to be able to deal with the new generation of computers, namely quantum computers. However, current software applications will not abruptly disappear or become useless. It is very likely

[1] Technical University of Munich, University of Antwerp & Flanders Make, Germany & Belgium. armin.moin@tum.de
[2] University of Antwerp & Flanders Make, Belgium. moharram.challenger@uantwerpen.be
[3] University of Reading, United Kingdom. atta.badii@reading.ac.uk
[4] Technical University of Munich & Munich Data Science Institute (MDSI), Germany. guennemann@in.tum.de

doi:10.18420/inf2022_95

that for a relatively long period of time, we will have to deal with a mixture of classical and quantum computers, algorithms, and data. Hence, the current challenges introduced by pervasive, distributed computing, which intrinsically involve heterogeneity, will even become more crucial, due to the new hardware technologies and architectures, which will require a fundamentally different paradigm and model for the programming and execution.

Based on our experiences with the heterogeneous and distributed Internet of Things (IoT) services and the highly complex systems of systems, namely the smart Cyber-Physical Systems (CPS), as well as AI [ML-20, MRG18, Mo20], where we found the applications of the Model-Driven Software Engineering (MDSE) paradigm, particularly the Domain-Specific Modeling (DSM) methodology [KT08] with full code generation beneficial, we maintain the idea of enabling MDSE for quantum computers and hybrid applications, that shall run on a mixture of quantum and classical computers. The overall idea of modeling quantum programs has been recently proposed independently by Delgado and Gonzalez [PDPG20], Ali and Yue [AY20], as well as Gemeinhardt et al. [GGW21].

The contribution of this paper is proposing our novel vision towards MDE4QAI, where our current open source technology stack concerning domain-specific MDSE for smart CPS and the IoT [ML-20] should be extended to support QAI, including Quantum ML (QML) and Quantum Multi-Agent Systems (QMAS). This inherently includes the hybrid quantum-classical applications as well.

The rest of this paper is structured as follows: Section 2 briefly reviews the state of the art. Further, we propose our vision concerning MDE4QAI with a focus on smart CPS and the IoT in Section 3. Finally, we conclude and suggest the future work in Section 4.

2 State of the Art

Today's non-quantum (i.e., *classical*) computers deal with bits. However, built based on quantum mechanics, quantum computers deal with quantum bits, known as *qubits*. A qubit may simultaneously represent a 0 and a 1, with certain degrees or probabilities of 0-ness and 1-ness. This property is called *superposition*. For example, if we have 2 bits, they can encode 4 possible states, namely 00, 01, 10, and 11, whereas 2 qubits can be represented through the following linear combination in the bra-ket (Dirac) notation: $\alpha_0 |00\rangle + \alpha_1 |01\rangle + \alpha_2 |10\rangle + \alpha_3 |11\rangle$, where α_i are complex numbers, called amplitudes, and their squares (i.e., $|\alpha_i|^2$) sum to 1. In fact, $|\alpha_i|^2$ is the probability that the quantum program will be in the state that is associated to the amplitude α_i, once the qubits are read by the quantum computer. In addition to superposition, there is another characteristic of quantum mechanics, thus quantum information/computation, called *entanglement*. If a pair or a group of quantum particles or qubits in our context are entangled, it is not possible to describe the quantum state of one particle/qubit independently of the other one(s). Therefore, the outcome of the measurement for one of the entangled particles/qubits

is always correlated with the outcome of the measurement for the other one(s) even if they are far away [AY20, NC11].

Quantum-encoded data require incomparably less storage space. Just as multiple bits can represent exponentially many numbers, a multi-qubit system can represent a superposition of exponentially many bit strings. *Quantum data* construct one pillar of QC, thus, one pillar of QAI/QML [Ar]. The second pillar is shaped by *quantum algorithms*. Note that any classical algorithm can be executed on quantum computers too. However, the QC community aims for exploiting the unique quantum properties, such as entanglement, in algorithms, to enable performance leap. The algorithms that inherently benefit from the essential properties of QC are called quantum algorithms. Nevertheless, the unsolvable computational problems (in terms of computability) on classical computers remain unsolvable on quantum computers too. Only the solvable problems could be solved much more efficiently on quantum processors. Examples include, but are not limited to the Shor's Algorithm for integer factorization and discrete logarithms [Sh94], the Grover's algorithm for efficient database search [Gr96], graph algorithms and random walks, as well as Principle Component Analysis (PCA) [CEea18].

There exist various models of computation for QC, such as quantum circuit/logic gate, adiabatic/annealing, topological, quantum walks, one clean qubit, measurement-based and quantum Turing machines [Jo08]. In practice, the quantum circuit model is well established, whereas quantum annealing is also deployed in certain implementations, for example, the Quantum Processing Units (QPU) of D-Wave Systems [DW]. In fact, it is already proven that any quantum circuit algorithm can be transformed into a quantum annealing algorithm with the exact same time complexity and vice-versa, hence showing they are essentially equivalent [YHW18, Jo]. Furthermore, the key technologies for the actual physical implementation of quantum processors comprise superconducting, ultra-cold atoms (e.g., trapped-ions), quantum optics and spin-based. For instance, Linke et al. [LMea17] conducted experiments on two 5-qubit quantum computers with the quantum circuit model, which had two different technologies/architectures, namely superconducting and trapped-ions. They illustrated that the superconducting system offered faster gate clock speeds and a solid-state platform, whereas the other one featured superior qubits and reconfigurable connections. They concluded that the performance of those systems reflected the topology of connections in the base hardware, thus supporting the idea that quantum software and hardware should be ideally co-designed [LMea17].

IBM [IQna] adopted the quantum circuit model of computation and the superconducting technology. Analogue to the low-level Assembly programming languages for classical computers, the OpenQASM [Cr17, Op] provided an intermediate representation for the quantum instructions of the IBM quantum computers. A higher level API is provided by the open source SDK, Qiskit [Qi]. The programmer may either create the circuits in Python through Jupyter notebooks or via a GUI. Additionally, Microsoft provided an open source SDK, called Q# [SRea18], which can be used with their Azure Quantum open cloud ecosystem [Az]. They enabled access to diverse quantum hardware technologies of

their partners, ranging from trapped-ions (Honeywell/IONQ) to superconducting (Quantum Circuits, Inc.). Further, D-Wave Systems [DW] offered the open source Ocean SDK [os] for using their superconducting quantum annealing QPUs. Last but not least, the Google quantum supremacy experiment [AAea19] (see Section 1) was conducted on a superconducting quantum computer, specifically a Noisy Intermediate-Scale Quantum (NISQ) [Pr18] processor with 53 qubits [TC]. Most of the current gate model solutions are NISQ.

Google/Alphabet, who provided TenforFlow [AAea15], an open source framework and Python library for ML using Artificial Neural Networks (ANN), particularly deep learning, are also one of the pioneers of QAI/QML. They extended TensorFlow for hybrid quantum-classical ML. This extended open source library is called TensorFlow Quantum (TFQ) [TC]. It deploys Cirq [Ci], an open source Python library for writing, manipulating, and optimizing quantum circuits and running them against the NISQ quantum computers and simulators. Cirq is the alternative solution of Google to Qiskit [Qi], Q# [SRea18] and Ocean [os]. The latter also offers special support for ML (e.g., for probabilistic models). However, QAI is not limited to QML. Already in 2004, Klusch [Kl04b, Kl04a] proposed the perspective of using Intelligent Agents (IA) and Multi-Agent Systems (MAS) in QC. Other works in the IA/MAS community (e.g., Chen et al. [CHH07]) also studied QC or specific properties of it. However, real-world implementations of IA/MAS on quantum computers, or more precisely hybrid quantum-classical MAS are emerging only recently, for example, see [Ne20, Ki20].

3 Proposed Vision: MDE4QAI

In this work, we focus on MDE in the context of Software Engineering (SE), thus MDSE. We first draw our proposed vision regarding MDE4QC and then concentrate specifically on MDE4QAI. The MDSE paradigm, which provides abstraction and automation can hide the complexity and increase the productivity of the software development. There exist many benefits in deploying software models as the central artifacts in the Software Development Life-Cycle (SDLC), for example, concerning the early validation, verification, analysis and simulation, as well as the automated generation of the implementation artifacts, such as the source code and the documentation. In our prior work, ML-Quadrat [MRG18, Mo20, ML-20], which was based on ThingML/HEADS [Thi16, Ha16, HE], we deployed the DSM methodology [KT08] of MDSE with full code generation to support the development of smart IoT services and smart CPS applications that needed to deploy ML. Thus, we provided a unified layer of abstraction to the practitioners, namely the users of the Domain-Specific Modeling Language (DSML), despite all of the heterogeneity of the underlying IoT platforms and the diversity of the ML libraries, frameworks and methods. Further, other examples in the literature include, but, are not limited to PIM4Agents [WK11], that abstracted from various MAS platforms (e.g., JACK, JADE, and Jadex), and the work of Challenger et al. [Ch14]. Since the Agent-Oriented Software Engineering (AOSE) paradigm fits very well to the IoT/CPS domain (see [PVHT18, GB14]), the integration of

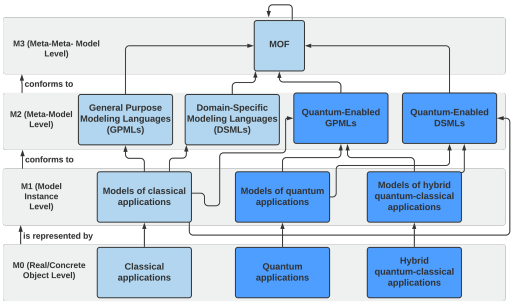


Fig. 1: The proposed vision of MDE4QC in the context of MOF-based meta-modeling.

the approaches as mentioned above (e.g., ML-Quadrat and PIM4Agents) constitutes the first part of the proposed vision in this paper. Note that the prior work on which ML-Quadart was based [MRG18, Mo20, ML-20], namely ThingML/HEADS [Thi16, Ha16, HE], did not support AI at all. However, ML-Quadrat enabled ML, which was the first step towards enabling AI and cognitive capabilities – which are vital for the smart services in the IoT-related domains – in the modeling framework of ThingML. However, extending this AI-support towards automated reasoning and intelligent agents will be the next necessary step in this direction.

In addition, the second part of the proposed vision is enabling classical modeling languages such that they can support the modeling of hybrid quantum-classical software applications. Figure 1 illustrates this idea in the context of the Object Management Group (OMG) and ISO/IEC standard for meta-models, called Meta-Object Facility (MOF) [OM19, IS05]. While we share the overall idea proposed by Gemeinhardt et al. [GGW21], we advocate the extension of the existing modeling languages, both the General-Purpose Modeling Languages (GPMLs), such as UML/SysML and the DSMLs, such as ML-Quadrat [MRG18, Mo20, ML-20] and PIM4Agents [WK11]. This vision is in line with the approach of Delgado and Gonzalez [PDPG20], who proposed a UML extension for QC, called Q-UML. In contrast, Gemeinhardt et al. [GGW21] considered a separate category for quantum modeling languages, next to GPMLs (UML) and DSMLs. However, we believe that no matter how far the programming paradigms, APIs and technologies of the classical and quantum computers are, modeling languages and DSMLs of the future shall continue to offer sufficiently extendable and adaptive models to support the requisite abstractions for the pure classical/quantum and the hybrid quantum-classical software systems. This is because we expect the co-existence of classical and quantum processors with various architectures and technologies in the foreseeable future.

Stepping into the world of quantum Information and Communication Technologies (ICT) and the hybrid quantum-classical ICT world, we shall expect not only more distributed computation, but also much more heterogeneity. This may comprise quantum/hybrid vs. classical data/information, quantum/hybrid vs. classical algorithms/computation, and

quantum vs. classical communication methods and technologies. Hence, a mixture of classical and quantum processors, each with different architectures and computational powers, as well as their individual constraints (e.g., in terms of the electric power consumption or the environmental requirements, such as, concerning the temperature, noise, and possible signal interference) for the coming years and possibly decades will be likely. In this setup, MDSE shall be capable of providing even more added value than before, given the unprecedented heterogeneity of the target hardware and software platforms, as well as the enormous complexity of the counter-intuitive and alien world of QC. MDSE will be capable of abstracting from the low-level APIs and programming paradigms, thus offering a higher level of abstraction, where practitioners can model their target applications, regardless of whether the applications shall run fully or partially on quantum processors of potentially diverse technologies and architectures. The advantages of MDSE and DSM may go far beyond the automated generation of the implementation artifacts. Model analysis for program comprehension and model checking for formal verification in the early design phases, which are typically interesting in the classical systems will be even more vital for QC and hybrid applications, due to the abundance of heterogeneous and distributed hardware and software platforms, the unfamiliarity of the technologies, as well as the limited access to the QC resources. The latter is currently addressed through the QC simulators on classical computers, but for the real-world distributed use-case scenarios with the hybrid quantum-classical applications, this does not suffice. However, the said vision implies the formalization and realization of the QC domain semantics in general, as well as the semantics and constraints of the specific target platforms for QC, ideally directly on the meta-model layer, but also possibly through the model-to-model and the model-to-code transformations. In other words, the domain knowledge regarding QC must be transferred from the human experts and the available libraries (mentioned in Section 2) to the modeling languages and tools.

Since our focus is on MDE4QAI, we plan to extend our existing DSML and modeling tool, ML-Quadrat towards QAI, specifically Quantum ML (QML) and Quantum MAS (QMAS). Currently, the practitioner using ML-Quadrat or similar domain-specific MDSE tools, does not need to be concerned about the technical details of the underlying platforms and their APIs or programming paradigms. In fact, in some cases, the practitioner might not even know at the design-time which specific IoT hardware/software platforms or communication protocols will be used at the run-time. Similarly, they might not have any idea at the design-time whether the ML algorithms will run on CPUs, GPUs, Tensor Processing Unit (TPUs) or a combination of them. Many such decisions can be left open to be made at the run-time, or can be set according to certain criteria. What is important is to offer the practitioner a higher layer of abstraction (i.e., the modeling level), where they can focus on the business logic without being concerned about the underlying technologies, which might change at a fast pace or become (un)available at short notice (see the Models@Runtime approaches, such as [FNea12]). By expanding the range of supported platforms and technologies to QC, we will consider not only CPUs, GPUs and TPUs, but also various QPUs with different models of computation, architectures, technologies, hence diverse powers and

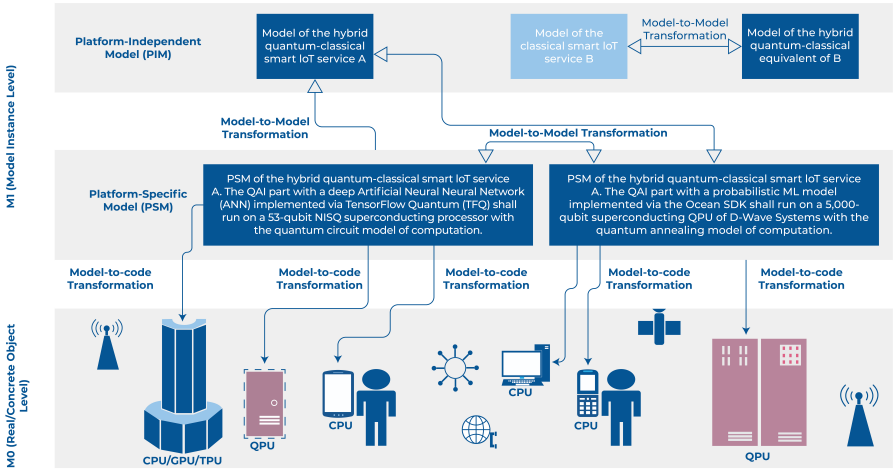


Fig. 2: The proposed vision of MDE4QAI in the context of Model-Driven Architecture.

constraints. The envisioned scenario for deploying MDE for QAI (MDE4QAI) in the context of Model-Driven Architecture (MDA) [OM14] is illustrated in Figure 2. Model-to-code transformations shall enable the fully-automated code generation. Moreover, it would be interesting to examine how far model-to-model transformations might be able to facilitate at the modeling level the following transformations: (i) A purely classical application to its hybrid/quantum equivalent or vice-versa, or (ii) a purely quantum application to its hybrid/classical equivalent or vice-versa, or (iii) a quantum/hybrid application designed for a certain model of computation, say quantum circuits to its equivalent designed for another model of computation, such as quantum annealing or vice-versa, or (iv) a quantum/hybrid application tuned for running on a specific quantum processor with a particular architecture and technology, say trapped-ions with X qubits to its equivalent optimized for running on another quantum processor with a different architecture and technology, such as a superconducting NISQ with Y qubits. Last but not least, as mentioned, analysis, verification and simulation in the early design stages at the modeling level shall be enabled.

4 Conclusion & Future Work

In this paper, we proposed our vision towards MDE4QAI, where the MDE/MDSE paradigm shall be used to facilitate the transition towards the world of QC and QAI. We advocate enhancing existing general purpose and domain-specific modeling languages, including our prior work for the IoT/CPS domain, ML-Quadrat [ML-20], in order to enable a hybrid quantum-classical model for the distributed and heterogeneous software systems, particularly the smart IoT/CPS services/applications of the near future, which shall run on an even more

diverse set of underlying technologies, including various quantum processors with different architectures and technologies.

Acknowledgment

The authors would like to thank Dalibor Hrg for sharing the insights.

Bibliography

[AAea15] TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, Software available from tensorflow.org.

[AAea19] Arute, Frank; Arya, Kunal; et al.: Quantum supremacy using a programmable superconducting processor. Nature, 574(7779):505–510, 2019.

[Ar] To understand quantum computing, start with quantum data. https://www.thomsonreuters.com/en-us/posts/news-and-media/understand-quantum-machine-learning-start-quantum-data/, Accessed: 2021-06-24.

[AY20] Ali, Shaukat; Yue, Tao: Modeling Quantum Programs: Challenges, Initial Results, and Research Directions. APEQS 2020, Association for Computing Machinery, New York, NY, USA, p. 14–21, 2020.

[Az] Azure Quantum. https://azure.microsoft.com/en-us/services/quantum/#product-overview, Accessed: 2021-06-28.

[CEea18] Coles, Patrick J.; Eidenbenz, Stephan J.; et al.: Quantum Algorithm Implementations for Beginners. CoRR, abs/1804.03719, 2018.

[Ch14] Challenger, Moharram; Demirkol, Sebla; Getir, Sinem; Mernik, Marjan; Kardas, Geylani; Kosar, Tomaž: On the use of a domain-specific modeling language in the development of multiagent systems. Engineering Applications of Artificial Intelligence, 28:111–141, 2014.

[CHH07] Chen, Kay-Yut; Hogg, Tad; Huberman, Bernardo A.: Behavior of Multi-Agent Protocols Using Quantum Entanglement. In: Quantum Interaction, Papers from the 2007 AAAI Spring Symposium, Technical Report SS-07-08, Stanford, California, USA, March 26-28, 2007. AAAI, pp. 1–8, 2007.

[Ci] Cirq. https://github.com/quantumlib/Cirq, Accessed: 2021-06-24.

[Cr17] Open Quantum Assembly Language, https://arxiv.org/pdf/1707.03429.pdf.

[DW] D-Wave Systems. https://www.dwavesys.com, Accessed: 2021-06-28.

[FNea12] Fouquet, François; Nain, Grégory; et al.: An Eclipse Modelling Framework Alternative to Meet the Models@Runtime Requirements. In: Model Driven Engineering Languages and Systems. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 87–101, 2012.

[GB14] Geisberger, Eva; Broy, Manfred, eds. Living in a networked world. Integrated research agenda Cyber-Physical Systems (agendaCPS). acatech STUDY. Herbert Utz Verlag, Munich, Germany, 2014.

[GGW21] Gemeinhardt, Felix; Garmendia, Antonio; Wimmer, Manuel: Towards Model-Driven Quantum Software Engineering. In: Second International Workshop on Quantum Software Engineering. ICSEW: Proceedings of the IEEE/ACM International Conference on Software Engineering Workshops, 2021.

[Gr96] Grover, Lov K.: A Fast Quantum Mechanical Algorithm for Database Search. In: Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing. STOC '96, Association for Computing Machinery, New York, NY, USA, p. 212–219, 1996.

[Ha16] Harrand, Nicolas; Fleurey, Franck; Morin, Brice; Husa, Knut Eilif: ThingML: A Language and Code Generation Framework for Heterogeneous Targets. In: Proceedings of the ACM/IEEE 19th International Conference on Model Driven Engineering Languages and Systems. MODELS '16, 2016.

[HE] HEADS. http://heads-project.eu, Accessed: 2020-09-08.

[IQna] IBM Quantum Computing. https://www.ibm.com/quantum-computing/.

[IS05] ISOMOF: ISO/IEC 19502:2005, Information technology — Meta Object Facility (MOF). Standard, ISO/IEC, 2005.

[Jo] Quantum adiabatic and quantum circuit algorithms are equivalent, say physicists. https://physicsworld.com/a/quantum-adiabatic-and-quantum-circuit-algorithms-are-equivalent-say-physicists/, Accessed: 2021-06-24.

[Jo08] Jordan, Stephen P.: Quantum Computation Beyond the Circuit Model. PhD thesis, Massachusetts Institute of Technology (MIT), 2008. https://arxiv.org/pdf/0809.2307.pdf.

[Ki20] Kirke, Alexis: Testing a hybrid hardware quantum multi-agent system architecture that utilizes the quantum speed advantage for interactive computer music. Journal of New Music Research, 49(3):209–230, 2020.

[Kl04a] Klusch, Matthias: Toward Intelligent Agents on Quantum Computers. In: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 3. AAMAS '04, IEEE Computer Society, USA, p. 1518–1519, 2004.

[Kl04b] Klusch, Matthias: Toward Quantum Computational Agents. In (Nickles, Matthias; Rovatsos, Michael; Weiss, Gerhard, eds): Agents and Computational Autonomy. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 170–186, 2004.

[KT08] Kelly, Steven; Tolvanen, Juha-Pekka: Domain-Specific Modeling: Enabling Full Code Generation. Wiley-IEEE Computer Society Pr, 1st edition, 2008.

[LMea17] Linke, Norbert M.; Maslov, Dmitri; et al.: Experimental comparison of two quantum computing architectures. Proceedings of the National Academy of Sciences, 114(13):3305–3310, 2017.

[ML-20] ML-Quadrat. https://github.com/arminmoin/ML-Quadrat, Accessed: 2020-09-12.

[Mo20] Moin, Armin; Rössler, Stephan; Sayih, Marouane; Günnemann, Stephan: From Things' Modeling Language (ThingML) to Things' Machine Learning (ThingML2). In: Proceedings of MODELS 2020 Satellite Events (Poster Companion/Extended Abstract). 2020.

[MRG18] Moin, Armin; Rössler, Stephan; Günnemann, Stephan: ThingML+: Augmenting Model-Driven Software Engineering for the Internet of Things with Machine Learning. In: Proceedings of the 2nd International Workshop on Model-Driven Engineering for the Internet of Things (MDE4IoT). 2018.

[NC11] Nielsen, Michael A.; Chuang, Isaac L.: Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press; 1st edition, UK, 2011.

[Ne20] Neumann, Niels M. P.; de Heer, Paolo B. U. L.; Chiscop, Irina; Phillipson, Frank: Multi-agent Reinforcement Learning Using Simulated Quantum Annealing. In: Computational Science – ICCS 2020. Springer International Publishing, Cham, pp. 562–575, 2020.

[OM14] OMGMDA: Model Driven Architecture (MDA), MDA Guide rev. 2.0. Standard OMG Document ormsc/14-06-01, Object Management Group, Boston, MA, USA, 2014.

[OM19] OMGMOF: OMG Meta Object Facility (MOF) Core Specification Version 2.5.1. Standard, Object Management Group, Inc. (OMG), 2019.

[Op] OpenQASM. https://github.com/qiskit/openqasm, Accessed: 2021-06-24.

[os] D-Wave Ocean SDK. https://github.com/dwavesystems/dwave-ocean-sdk, Accessed: 2021-06-28.

[PDPG20] Pérez-Delgado, Carlos A.; Perez-Gonzalez, Hector G.: Towards a Quantum Software Modeling Language. In: Proc. of the IEEE/ACM 42nd Int. Conf. on Software Engineering Workshops. ICSEW'20, Association for Computing Machinery, New York, NY, USA, p. 442–444, 2020.

[Pr18] Preskill, John: Quantum Computing in the NISQ era and beyond. Quantum, 2:79, August 2018.

[PVHT18] Pico-Valencia, Pablo; Holgado-Terriza, Juan: Agentification of the Internet of Things: A systematic literature review. International Journal of Distributed Sensor Networks, 14(10):1550147718805945, 2018.

[Qi] Qiskit. https://qiskit.org, Accessed: 2021-06-24.

[Sh94] Shor, P.W.: Algorithms for quantum computation: discrete logarithms and factoring. In: Proceedings 35th Annual Symposium on Foundations of Computer Science. pp. 124–134, 1994.

[SRea18] Svore, Krysta; Roetteler, Martin; et al.: Q#. Proceedings of the Real World Domain Specific Languages Workshop 2018 on - RWDSL2018, 2018.

[TC] TensorFlow Quantum (TFQ): Quantum machine learning concepts. https://www.tensorflow.org/quantum/concepts, Accessed: 2021-06-24.

[Thi16] ThingML. https://github.com/TelluIoT/ThingML, Accessed: 2020-04-29.

[WK11] Warwas, Stefan; Klusch, Matthias: Making Multiagent System Designs Reusable: A Model-Driven Approach. In: Proceedings of the International Conferences on Web Intelligence and Intelligent Agent Technology. WI-IAT '11, IEEE Computer Society, USA, p. 101–108, 2011.

[YHW18] Yu, Hongye; Huang, Yuliang; Wu, Biao: Exact Equivalence between Quantum Adiabatic Algorithm and Quantum Circuit Algorithm. Chinese Physics Letters, 35:110303, 10 2018.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1133

Implementations for Shor's algorithm for the DLP

Alexander Mandl[1], Uwe Egly[2]

Abstract: Shor's algorithm for solving the discrete logarithm problem is one of the most celebrated works in quantum computing. It builds upon a quantum circuit performing modular exponentiation. As this is a comparatively expensive process, many approaches for reducing both the number of used qubits and the number of applied gate operations have been proposed. We provide quantum circuits in Qiskit for three different implementation proposals aiming to reduce space complexity and compare their performance regarding their asymptotic gate complexity. We make use of the circuit implementations and Qiskit's simulation capabilities to compare the actual number of applied gate operations in compiled circuits for small problem instances to aid future applications of this algorithm.

Keywords: quantum computing; discrete logarithm; implementation; Shor's algorithm

1 Introduction

Shor's polynomial-time algorithms for prime factorization and the discrete logarithm problem (DLP) [Sh97] are among the most groundbreaking results in the field of quantum computation. These problems are of particular interest since their hardness is the basis of the security of various cryptographic systems, because the problems are believed to not be solvable on a classical computer in polynomial time. Many variants and reformulations [ME98, PZ03, Ka17] of Shor's algorithms have since been studied and, although the physical realization of quantum computers powerful enough to execute them for meaningful inputs still lies in the future, the field of quantum computation has grown considerably.

Shor's algorithms make use of the fact that the quantum Fourier transform (QFT) can be used to estimate the eigenvalues of quantum transformations in a process referred to as quantum phase estimation (QPE) [ME98]. This operation uses a quantum oracle which computes some unitary operation U which is known to have the eigenvector $|\psi\rangle$ and the corresponding eigenvalue $e^{2\pi i\phi}$. The output of the overall process is then the closest integer to $\phi 2^n$ for n depending on the quantum circuit size. Furthermore, it can be shown that this value is obtained with a probability of at least $4/\pi^2$ [ME98].

We focus on the DLP which aims to find an integer m such that $g^m \equiv b \bmod p$ for a given integer-valued generator g, an integer b, and a prime number p. The quantum oracle

[1] Institute of Architecture of Application Systems, University of Stuttgart, Germany, alexander.mandl@iaas.uni-stuttgart.de
[2] Institute of Logic and Computation, TU Wien, Austria, uwe.egly@tuwien.ac.at

doi:10.18420/inf2022_96

involved in the QPE process is a quantum circuit U_c that performs the modular multiplication $U_c |y\rangle = |cy \bmod p\rangle$.

Since the QPE algorithm requires the repeated controlled application of this oracle for different constants, which amounts to the exponentiation $|x\rangle U_{c^x} |y\rangle = |x\rangle |c^x y \bmod p\rangle$, it is usually referred to as the *modular exponentiation* oracle. In the description of this circuit it is assumed that the constant c is built into the circuit.

Given the multiplicative order of the generator g modulo p, described by $r = |\langle g \rangle|$, the eigenvalues of this transformation have the form

$$U_b |\psi_k\rangle = e^{\frac{2\pi i k m}{r}} |\psi_k\rangle ,$$

for $0 \leq k < r$, with eigenvectors $|\psi_k\rangle$. Herein, b refers to the input of the DLP from above.

The algorithm uses QPE to estimate the eigenvectors for both the oracle U_b performing the multiplication with b modulo p and U_g which performs the same operation with the integer g and has eigenvalues $\exp(2\pi i k/r)$. This leads to estimates of the fractions $(km \bmod r)/r$ as well as k/r, which can be used to infer the discrete logarithm m on a classical computer.

The fundamental result of Shor's work is that the best approximations to these two fractions are obtained with a constant probability greater than zero, which implies that the quantum circuit provides the correct measurements after a constant number of repetitions. For the purpose of inferring m it has to hold that the measured k and the order r have to be coprime for the algorithm to succeed. However, for large r it was shown that a lower bound for this probability can be given [Sh97], which results in a total runtime for the whole algorithm of

$$O((\log \log p) T(n)), \tag{1}$$

where $T(n)$ describes the number of gate operations required to perform the oracle operation.

Although Shor's algorithm succeeds in solving the DLP in a theoretical setting, the application of this process using real machines still remains an open problem. Currently, there are no quantum computers powerful enough to run this algorithm even for relatively small problem instances of a few bits. This stems from the fact that the involved transformations have to use a number of extra temporary qubits, so-called *ancilla* qubits, which greatly increase the resource demand of the whole algorithm beyond the capabilities of most current quantum computers. On the other hand, the sheer number of operations both in the modular exponentiation oracle as well as in the overall algorithm introduces the possibility for errors in the physical realization of the quantum transformations, which leads to unusable measurement results even for moderately small instances. Both of these obstacles are further highlighted by the algorithm descriptions in the later parts of this text.

However, the steady increase in processing power of quantum computers might lead to future machines powerful enough to run the algorithm. To aid future realizations of Shor's algorithm, this work provides implementations[3] of the algorithm for the discrete logarithm

[3] The source code can be found at `https://github.com/mhinkie/ShorDiscreteLog`

problem in Qiskit[4] that focus on conserving resources both in the number of qubits and gate operations that are applied.

As described above, the modular exponentiation oracles U_{b^x} and U_{g^x} are integral parts of this quantum algorithm. Therefore, we examine efficient implementation proposals for these arithmetic operations. Implementation variants of modular multiplication on a quantum computer range from repeated addition [Be03, HRS17] to more involved approaches such as Montgomery modular multiplication or Barrett reduction [RC18]. We implement and compare three different proposals for performing this arithmetic process and present a comparison regarding both, their asymptotic gate complexity as well as regarding the actual number of applied gate operations for small examples. Furthermore, there are multiple variants of the overall phase estimation approach that save qubits by restructuring the circuit [Be03, ME98], some of which were also implemented in Qiskit for this work (see [Ma21] for details). We selected this particular set of algorithms as they focus on minimizing the space complexity of the modular exponentiation operation. This allows us to simulate the circuits for larger inputs on classical devices and will also allow users to execute the overall algorithm in the future on smaller quantum computers.

The comparison presented in Section 3 has to take into account that a decrease in space complexity, i. e. in the number of qubits, often goes hand-in-hand with an increase in time complexity and vice versa. For this reason, we not only compare the number of used qubits but also the gate complexity for large n and for small examples using compiled circuits. Of course, the actual computing time and space complexity are also influenced by a number of other factors that depend on the hardware that will execute the algorithm in the future such as the error correction routines. Due to their dependence on the actual physical implementation these factors were not studied explicitly in this work. Gidney and Ekerå [GE21] analyze these requirements and give an approximation of the number of qubits required for an error corrected implementation of Shor's algorithm on noisy systems.

2 Modular exponentiation

The modular exponentiation operation is the most costly component of the overall algorithm, as such implementations usually focus on keeping the number of required operations as small as possible. Shor notes in [Sh97] that the Schönhage-Strassen multiplication algorithm would theoretically be the fastest one for this problem for large input sizes. However, implementations of this algorithm as a quantum circuit require a significant amount of temporary storage space [Za98, RC18] which is why we do not consider this algorithm.

The most straightforward way of performing modular multiplication is by simply repeating a modular addition circuit using the binary decomposition of the quantum input integer x:

$$c \cdot x \bmod p = c2^0 \cdot x_1 + c2^1 \cdot x_2 + \cdots + c2^{n-1} \cdot x_n \bmod p. \tag{2}$$

[4] https://qiskit.org/

Therefore, before explaining the multiplication algorithm, different ways of performing modular addition and in turn modular subtraction are investigated.

2.1 Addition using the QFT

Draper [Dr00] proposes a quantum circuit that adds two n qubit integers by using the fact that the state after the QFT of some integer a is separable and can be described as

$$\bigotimes_{k=1}^{n} |\phi_k(a)\rangle \text{ , where } |\phi_k(a)\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + \exp\left(2\pi i \frac{a}{2^k}\right) |1\rangle \right).$$

The addition circuit itself applies Qiskit's phase gate P. This operation maps arbitrary states $\alpha |0\rangle + \beta |1\rangle$ to $\alpha |0\rangle + \exp(i\lambda)\beta |1\rangle$ for a real value λ supplied at the time of construction. Specifically, using $\lambda = \frac{2\pi b}{2^k}$ for a known integer b, it maps one Fourier-transformed qubit to

$$P\left(\frac{2\pi b}{2^k}\right) |\phi_k(a)\rangle = \frac{1}{\sqrt{2}} \left(|0\rangle + \exp\left(2\pi i \frac{a+b}{2^k}\right) |1\rangle \right) = |\phi_k(a+b)\rangle .$$

Performing this process for each qubit individually will give the Fourier-transformed result of the addition in linear time, assuming the input is already supplied in its Fourier-transformed representation. This circuit can be extended to perform modular addition and modular multiplication as presented by Beauregard [Be03] and in the later sections of this text.

2.2 Bitwise addition

Another way of performing arithmetic on a quantum computer is to perform bitwise manipulations of the input similar to the processes on a classical computer. There are various descriptions of bitwise addition gates on a quantum computer ([RC18] gives a good overview). We chose to examine the implementations suggested by Häner et al. [HRS17] further. Although this bitwise addition gate uses comparatively many ancilla qubits, it is able to perform the calculations using *borrowed* ancilla qubits. These are idle qubits that can start in any arbitrary state and will be returned to the same state after the computation. Since in the overall multiplication circuit many such idle qubits are present, the number of used qubits for the algorithm can be decreased this way.

As its core component, the circuit relies on a gate that calculates the final carry of an addition of two n bit integers a and b, where the latter is known at the time of construction of the circuit. Assuming the binary representation of the $n+1$ bit addition result $r = a + b$ is given as $r_{n+1} r_n \dots r_2 r_1$, this operation can be described as

$$CARRY(b) |a\rangle |g\rangle |0\rangle = |a\rangle |g\rangle |r_{n+1}\rangle .$$

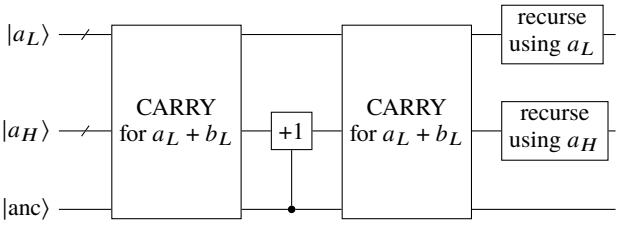


Fig. 1: Recursive addition circuit as presented in [HRS17] for a clean ancilla qubit. The adaptions required for borrowed ancilla qubits are described in [Ma21, HRS17].

The circuit needs an additional register $|g\rangle$ of size $n - 1$, which holds the borrowed qubits that are used as temporary storage during computation. How this circuit is constructed is presented in [HRS17] and details of the implementation of this process in Qiskit are presented in [Ma21]. The key concepts employed in this process will be explained here.

Using *CARRY* alone, a simple addition circuit could be constructed by iteratively computing the carry values of the addition of increasingly smaller subsets of the input with the supplied constant and toggling the resulting qubits accordingly. However, this approach would result in a circuit with $O(n^2)$ runtime [HRS17]. Häner et al. propose a different recursive approach: The n qubit input a is split into two approximately equal-sized subsets a_L and a_H with size $\lceil \frac{n}{2} \rceil$ and $\lfloor \frac{n}{2} \rfloor$ respectively. The resulting carry of the addition of a_L and the corresponding bits of the constant b_L is saved in another borrowed qubit and used to control an incrementer operation on the other half a_H of the input. The circuit then recursively computes the result for both subsets a_H and a_L to finally obtain the addition result. Figure 1 shows the corresponding quantum circuit.

The complete construction of the incrementer circuit used in each recursive step is presented in [HRS17, Ma21]. It is based on a quantum-quantum addition circuit presented in [TTK10].

The major difference between borrowed qubits as they are employed here and the zero-initialised ancilla qubits that are used in other parts of the algorithm is the fact that they can start in an arbitrary unknown state. To be able to still use these qubits to store intermediate carry values which control the application of other gates, Häner et al. reorder the operations such that they are applied before the control qubit is computed and reversed should the borrowed qubit indicate that the operation should not have been performed. To put it into other words, the operation is conditioned on the fact that this borrowed qubit is toggled.

2.3 Extension to modular addition

The modular exponentiation oracle requires a subroutine that not simply performs addition but performs addition modulo a prime module p. Using a zero-initialised ancilla qubit, both addition circuits presented here can be extended to perform modular addition by comparing the result of the addition $a + b$ with the module p.

In the bitwise adder presented in Subsection 2.2, the comparison is performed by using the *CARRY* gate to obtain the final carry of the addition $a + (b - p)$, which, as long as the register is appropriately sized, is 1 if and only if $a + b < p$. For the QFT-based adder (see Subsection 2.1), a subtraction of the module p is performed on the Fourier-transformed result before applying the inverse QFT to extract the most significant bit to decide if the subtraction by p has to be reversed.

The details of the implementation of this process in Qiskit (as well as the required uncomputation steps) can be found in [Ma21].

2.4 In-place modular multiplication

By employing the binary decomposition of the input x as presented in Equation (2), we construct a modular multiplier. This approach is described in [Be03] and similarly in [HRS17]. The circuit should compute the result $|cx \bmod p\rangle$ in the same register as the input was stored, while minimizing the number of ancillary qubits used.

The multiplication circuit is based on a circuit that performs the out-of-place multiplication $|x, y\rangle \rightarrow |x, (y + cx) \bmod p\rangle$ by repeatedly adding constants of the form $c \cdot 2^i \bmod p$ controlled by the individual input qubits $|x_i\rangle$. First the input $|x, 0\rangle$ is taken to $|x, cx \bmod p\rangle$ using the first out-of-place multiplication. After swapping both registers to obtain $|cx \bmod p\rangle\,|x\rangle$, the result will then remain in the leftmost register. The other register can be reset to the $|0\rangle$ state by performing another out-of-place multiplication with the constant $-c^{-1} \bmod p$ as the following equality shows:

$$x + (-c^{-1} \bmod p)(cx \bmod p) \bmod p = x - c^{-1}cx \bmod p = 0.$$

2.5 Montgomery modular multiplication

The quantum circuits proposed until now calculate the result of modular multiplication because the intermediate results are reduced modulo the prime p after each addition. This ensures that the number of required qubits is kept low since after each reduction the result is in the range $0 \leq x < p$. However, it also introduces unwanted overhead: The result has to be modified after each step. This overhead is especially noticeable in the case of Fourier space addition, where determining whether the result has to be reduced modulo p requires a full quantum Fourier transform and its inverse with runtime $O(n^2)$. In this section, we present the multiplication algorithm proposed by Rines and Chuang [RC18], which performs Montgomery modular multiplication [Mo85] on a quantum computer to address this problem.

As this algorithm often rearranges qubits between registers, we will highlight the size i of certain registers as $|x\rangle_i$ in the remainder of this section.

A naive optimisation of the multiplication algorithm is to use an appropriately sized register and perform all required additions on this register without ever transforming back to the computational basis. However, this implies that a division with p has to be performed at the end of the multiplication algorithm to obtain the result $|cx \bmod p\rangle$. Unfortunately, such a division requires a series of trial subtractions that are controlled by the most significant qubit of the value in the register and the extraction of this qubit again introduces a series of costly QFT operations.

Instead, the algorithm presented in [RC18] multiplies with the *Montgomery form* of the constant c, given by $c2^n \bmod p$. Afterward, the *Montgomery reduction* transforms the multiplication result $x \cdot (c2^n \bmod p)$ to $x \cdot (c2^n \bmod p)2^{-n} \bmod p$ to give the sought after modular multiplication result. The advantage of the Montgomery reduction is the fact that it does not require examinations of the most significant qubit but instead of the least significant qubit, which can be obtained in constant time by noting that

$$|\phi_1(a)\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle + (-1)^{a \bmod 2}|1\rangle\right) = H|a \bmod 2\rangle\,.$$

Here, $|\phi_1(a)\rangle$ represents the least significant qubit in the Fourier transformed representation of some integer a. Therefore, a simple Hadamard gate suffices to extract this qubit in the computational basis.

In the remainder of this subsection, we give an overview of the characteristics of the quantum circuit implementing Montgomery modular multiplication. The exact specification and further information on its implementation in Qiskit can be found in [Ma21].

To distinguish the content of a whole register in the computational basis and their Fourier-transformed counterpart, the former is referred to as $|a\rangle$ and the latter is referred to as $|\phi(a)\rangle$ for some integer a.

The circuit first performs an initial multiplication of the input with the Montgomery form of the constant. This multiplication is performed by repeated addition of the Fourier-transformed register content similar to Section 2.1 and is described as

$$|x\rangle_n|0\rangle_{2n+1} \rightarrow |x\rangle_n|\phi(x \cdot (c2^n \bmod p))\rangle_{2n+1}\,.$$

2.5.1 Montgomery reduction

Given the result of the multiplication as $t = x \cdot (c2^n \bmod p)$, the *estimation stage* of the circuit first computes the values $u = tp^{-1} \bmod 2^n$ and $s = (t - up)/2^n$. It can be shown that s is then congruent to and within p of the desired result $t2^{-n} \bmod p$ [Ma21]. Therefore only one correcting addition is needed at the end of this process to obtain the result.

In each step of the quantum circuit for the estimation stage, the currently least significant qubit of the register containing $|\phi(t)\rangle$ is removed and assigned as the new most significant

qubit of the register containing $|u\rangle$, which implicitly divides $|\phi(t)\rangle$ by two. This qubit is then used to control a subtraction by $p/2$. Therefore, when $|\phi(t')\rangle$ is the content of the first register at the beginning of this iteration, the ith iteration transforms $|\phi(t')\rangle_l |u\rangle_{i-1}$ to

$$\left|\phi\left(\frac{t'}{2} - \frac{u_i \cdot p}{2}\right)\right\rangle_{l-1} |u\rangle_i ,$$

with l such that $l + (i-1) = 2n + 1$.

After n repetitions of this process, the system reaches the desired state $|\phi(s)\rangle_{n+1} |u\rangle_n$ [RC18, Ma21]. Should the resulting value s be negative, the *correction stage* will afterwards add p to obtain the reduction result. This can be achieved by performing an addition conditioned on the $n+1$st qubit containing the sign of the result. After this correction, the system is in the state $|s_{n+1}\rangle |\phi(xc \bmod p)\rangle_n |u\rangle_n$.

It finally only remains to uncompute the values $|s_{n+1}\rangle$ and $|u\rangle$. For this task, Rines and Chuang [RC18] propose restoring the least significant bit of $|s\rangle$ in the register that holds the sign qubit $|s_{n+1}\rangle$. This can be achieved by performing a *CNOT* operation conditioned on the least significant qubit of $|\phi(xc \bmod p)\rangle$. To illustrate why this is the case, first assume $|s_{n+1}\rangle = |0\rangle$ and therefore no correction with p has been performed previously. Here, the *CNOT* operation simply copies the basis state present in the least significant qubit of the result, which was also the least significant qubit of $|s\rangle$ since no correction was performed. In the case when $|s_{n+1}\rangle = |1\rangle$ an addition with the odd integer p is performed, which causes the least significant qubit to flip. However, since $|s_{n+1}\rangle$ starts in the flipped state, the *CNOT* operation also suffices.

This new qubit is again treated as the new most significant qubit of $|u\rangle$ as was done during the estimation stage. Therefore it only remains to uncompute $|u\rangle_{n+1} = |tp^{-1} \bmod 2^{n+1}\rangle_{n+1}$. Since t and in turn u can be expressed using the multiplication that was performed at the beginning of the algorithm as

$$u = \sum_{i=0}^{n-1} x_i (c2^n \cdot 2^i \bmod p) p^{-1} \bmod 2^{n+1},$$

it is apparent that this register can be restored to $|0\rangle$ by performing n conditional subtractions with $(c2^n \cdot 2^i \bmod p)p^{-1}$. By extensions similar to the ones presented in Section 2.4 this circuit can further be used to perform in-place multiplication.

The main improvement in the depth of this circuit stems from the fact that the relatively expensive QFT only has to be performed at the beginning and at the end of the circuit (which in some instances can also be omitted) and at the beginning of the correction stage when the most significant qubit is extracted. This contrasts the approach presented in previous sections, where the QFT is repeated for every modular addition that is performed.

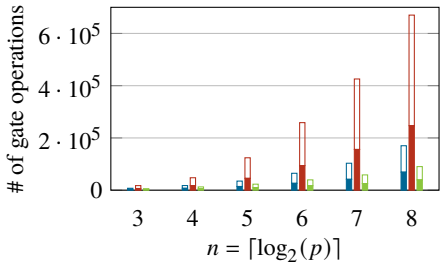


Fig. 2: The number of gate operations applied for each of the implementations presented in Section 2 for different input sizes n. The solid bars represent the share of *CNOT* operations in the total gate count. The QFT-based implementation by Beauregard [Be03] is shown in blue, the bitwise adder-based implementation by Häner et al. [HRS17] is shown in red and the Montgomery multiplication by Rines and Chuang [RC18] is shown in green.

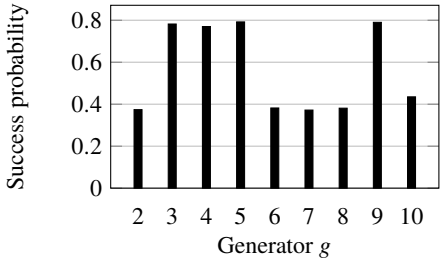


Fig. 3: The average success probability for multiple inputs to the full algorithm for the DLP for different generators and the prime $p = 11$. The upper bound for the obtainable success probability of $\phi(r)/r$ with $r = |\langle g \rangle|$ is clearly visible.

3 Results and comparison

We implemented the different circuits for modular exponentiation as well as optimisations for the overall phase estimation algorithm as they are presented in [Ma21] in Qiskit. The circuits and the overall implementation are available online[5].

Modular exponentiation gate	# of qubits	Gate operations $T(n)$
Beauregard [Be03]	$2n + 2$	$O(n^4)$
Häner, Roetteler and Svore [HRS17]	$2n + 1$	$O(n^3 \log(n))$
Rines and Chuang [RC18]	$3n + 1$	$O(n^3)$

Tab. 1: The number of required qubits excluding control qubits and the gate complexity for the modular exponentiation circuits implemented in [Ma21].

The complete algorithm for computing the discrete logarithm has the asymptotic gate complexity given in Equation (1), with $T(n)$ as listed in Table 1.

First, we compare the asymptotic gate complexity of the different modular exponentiation circuits presented in Section 2. Table 1 shows that the circuit based on Montgomery modular multiplication should perform best with respect to the other implemented circuits. It does, however, also use the most ancilla qubits, which is highlighted by its increased space complexity in Table 1.

[5] `https://github.com/mhinkie/ShorDiscreteLog`

Using the concrete implementations of these circuits, it is now also possible to compare the actual number of operations that are applied for small examples (see Figure 2). To obtain a realistic comparison for current quantum computer architectures, we compile the circuits to the gate set $\{RZ, I, \sqrt{X}, X, CNOT\}$ as this is currently used in IBM's quantum systems[6]. Furthermore, we compile for the unrestricted qubit layout of Qiskit's simulator to not introduce architecture-dependent swap operations arising from varying physical qubit layouts.

Using our implementations it is now possible to calculate discrete logarithms on quantum computers supporting Qiskit's backend. The noise-free results for one specific example are shown in Figure 3 and highlight the dependence of the success probability on the order r of the generator.

4 Summary and conclusion

In addition to the experimental results, our circuit implementations and the code to execute example instances on a simulator are available publicly to enable further research. The comparison of these gates shows that although the bitwise addition circuit by Häner et al. [HRS17] outperforms the QFT-based multiplication circuit by Beauregard [Be03] for small examples, the latter is preferable since the QFT cost is not that significant for small n. Of the three studied implementations, the most efficient in the number of gate operations for large and for small n is the circuit presented by Rines and Chuang [RC18]. Its only drawback is the comparatively large number of required qubits. In future work, the implementations provided in this work can be used on real quantum systems to further examine Shor's algorithm, its applicability to the DLP and further examine the impact of gate errors on the measurement results.

Bibliography

[Be03] Beauregard, S.: Circuit for Shor's algorithm using 2n+3 qubits. Quantum Inf. Comput., 3(2):175–185, May 2003.

[Dr00] Draper, T. G.: Addition on a Quantum Computer. arXiv preprint quant-ph/0008033, August 2000.

[GE21] Gidney, C.; Ekerå, M.: How to factor 2048 bit RSA integers in 8 hours using 20 million noisy qubits. Quantum, 5:433, April 2021.

[HRS17] Häner, T.; Roetteler, M.; Svore, K. M.: Factoring using 2n+2 qubits with Toffoli based modular multiplication. Quantum Inf. Comput., 17(7 & 8), June 2017.

[Ka17] Kaliski Jr., B. S.: A Quantum "Magic Box" for the Discrete Logarithm Problem. Cryptology ePrint Archive, Report 2017/745, 2017. https://ia.cr/2017/745.

[6] https://quantum-computing.ibm.com/services?services=systems&system=ibmq_santiago

[Ma21] Mandl, A.: Quantum Algorithms for the Discrete Logarithm Problem. Master's thesis, TU Wien, 2021.

[ME98] Mosca, M.; Ekert, A.: The Hidden Subgroup Problem and Eigenvalue Estimation on a Quantum Computer. In: QCQC'98, Selected Papers. volume 1509 of LNCS. Springer, pp. 174–188, 1998.

[Mo85] Montgomery, P. L.: Modular multiplication without trial division. Mathematics of Computation, 44:519–521, 1985.

[PZ03] Proos, J.; Zalka, C.: Shor's discrete logarithm quantum algorithm for elliptic curves. Quantum Inf. Comput., 3(4):317–344, January 2003.

[RC18] Rines, R.; Chuang, I.: High Performance Quantum Modular Multipliers. arXiv preprint arXiv:1801.01081, January 2018.

[Sh97] Shor, P. W.: Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. SIAM J. Comp., 26(5), October 1997.

[TTK10] Takahashi, Y.; Tani, S.; Kunihiro, N.: Quantum Addition Circuits and Unbounded Fan-Out. Quantum Inf. Comput., 10(9):872–890, September 2010.

[Za98] Zalka, C.: Fast versions of Shor's quantum factoring algorithm. arXiv preprint quant-ph/9806084, 1998.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1145

Real-world application benchmark for QAOA algorithm for an electromobility use case

Marika Federer[1] Daniel Müssig [2] Steve Lenk[3] Jörg Lässig[4]

Abstract: To reduce CO_2 emissions in the mobility sector, battery electric service vehicles might play an important role in the future. Here, an optimal charging scheduling use case will be presented which includes local solar power generation for minimizing the power grid usage for electric service vehicles. Different formulations of the use case are given to illustrate the differences for classical and quantum-based optimization using a mixed integer linear program and a quadratic unconstrained binary optimization program, respectively. Addtionally, we study the complexity of our benchmark experiments by characterizing the respective QUBO matrices and the optimization landscapes. It is shown how the setting of the parameters of a certain experiment and its penalty function influences the complexity for a quantum-based optimizer. Additionally, we present a comparison of the computing times and summarize the current state of gate-based quantum computing for electromobility.

Keywords: QAOA; Electromobility; Quantum Computing; Energy

1 Introduction

Reducing greenhouse gases dramatically is one of the major challenges in the near-term future. To achieve this goal within the mobility sector, by assuming that the number of electric vehicles worldwide registered will rise to 140 million or more by 2030 [ABG+20], it is obvious that the mobility sector might have a large impact in that challenge. However, optimal utility of fluctuating renewable energy generation is a requirement to achieve CO_2-neutral charging of electric vehicles. Next to the private sector, the service sector has to reduce its emissions, here we investigate the optimization of a charging scheduling problem for electric service vehicles of an airport. Based on the proposed progress of quantum computing for optimization applications, we present an approach which can be used for gate-based quantum computing and quantum annealing.

2 Model formalism and benchmark

We consider a system at the airport Erfurt-Weimar to charge a couple of electric service vehicles based on the flight schedule and the availability of photovoltaic (PV) electricity

[1] Fraunhofer IOSB-AST, Dep. Cognitive Energy Systems, Ilmenau, Germany marika.federer@iosb-ast.fraunhofer.de
[2] Fraunhofer IOSB-AST, Dep. Cognitive Energy Systems, Görlitz, Germany daniel.muessig@iosb-ast.fraunhofer.de
[3] Fraunhofer IOSB-AST, Dep. Cognitive Energy Systems, Ilmenau, Germany steve.lenk@iosb-ast.fraunhofer.de
[4] University of Applied Sciences Zittau/Görlitz, Brückenstraße 1, 02826 Görlitz, Germany jlaessig@hszg.de

doi:10.18420/inf2022_97

generation.
The target is to optimize the charging schedule for several service electric vehicles (EVs) in accordance with the flight schedule by using predominantly the electricity generated by a roof-top photovoltaic station to minimize the electricity costs and the CO_2 emissions.
The classical optimization problem can be formulated using unit-less values as an integer linear problem

$$z = min\left(\sum_t |j(t) - pv(t)|\right) \tag{1}$$

$$\text{w. r. t.} \quad j(t) \leq j_{max}(t) \tag{2}$$

$$\sum_t j(t) \geq e_{min} \tag{3}$$

$$\sum_t j(t) \leq e_{max}\,. \tag{4}$$

with the PV power forecast time series $pv(t)$, the (aggregated) maximal charging power forecast $j_{max}(t)$, the minimal and maximal charging energies e_{min} and e_{max}. Eq. (1) is our target (or cost) function of the integer linear program.
However, for the use of quantum computing, i. e., quantum annealing and quantum approximate optimization algorithm (QAOA) for gate-based quantum computing, the optimization problem is usually written as a quadratic unconstrained binary optimization (QUBO) problem [Ko14; Lu14]. Therefore, we replace the absolute value function of the target function by a square function. This step causes that certain entangled states of the original expression might be shifted to higher target values. However, these states contain higher charging power peaks, which might negatively affect the electric grid and, thus, should be generally avoided, i. e., this shift does not alter our optimization problem negatively.
The constraints have to be included into the target function, which can be done using quadratic penalty functions and penalty factors.
To yield the QUBO representation, the integer values $j(t)$ are mapped to an appropriate binary representation [Kü21; Lu14]

$$j(t) = \sum_{n=0}^{b(t)-2} 2^n j_n + (j_{max}(t) - 2^{b(t)-1} + 1)\, j_{b(t)-1}\,. \tag{5}$$

with the necessary number of binary variables $b(t) := \lceil \log_2(j_{max}(t) + 1) \rceil$. This choice of binary representation already ensures the first constraint (2) and, thus, only two penalty functions have to be considered.
The QUBO formulation of the optimization model representing the target function in eq. (1) as well as the two constraints of eqs. (3)-(4) can be written as

$$z = min\left(\sum_t (j(t) - pv(t))^2 + C_a + C_b\right) \tag{6}$$

where

$$C_a = P_a \left(\sum_t j(t) - t - e_{min} \right)^2 \tag{7}$$

$$C_b = P_b \left(\sum_t j(t) + v - e_{max} \right)^2 . \tag{8}$$

To obtain the quantum mechanical Hamiltonian [Lu14], the binaries are substituted in eq. (5) by

$$j_n = \frac{1}{2} \left(1 + \sigma_n^z \right) , \tag{9}$$

to act on respective qubits.
To be able to compare classical solvers and the hybrid quantum classical algorithm QAOA [FGG14], we created a benchmark consisting of 28 experiments with different complexity and varying number of binaries/qubits ranging from 3 to 23. The amount of qubits is limited by the quantum hardware. As an example, the experiment 28 of our benchmark is visualized in Fig. 1. The parameters can be found in Tab. 1

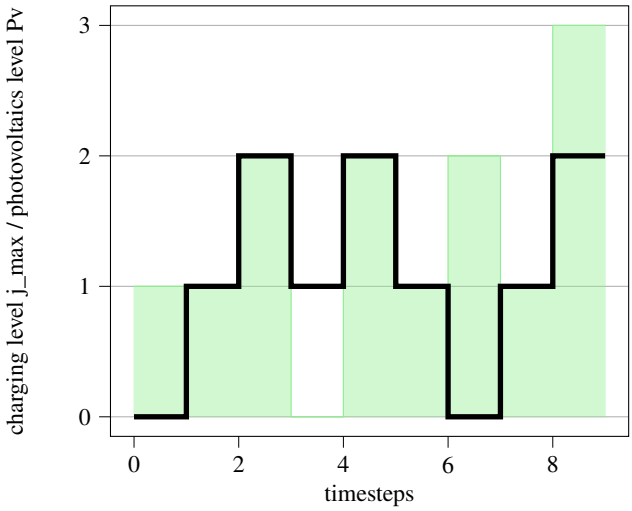


Fig. 1: Visualization of the input time series for experiment 28. The PV forecast time series $pv(t)$ is given in green, whereas the maximal charging power forecast $j_{max}(t)$ is marked by the bold black line.

3 Complexity

The details for four specific benchmark experiments are given in Tab. 1 to represent our complexity study. It can be seen from table that the number of time steps is increased

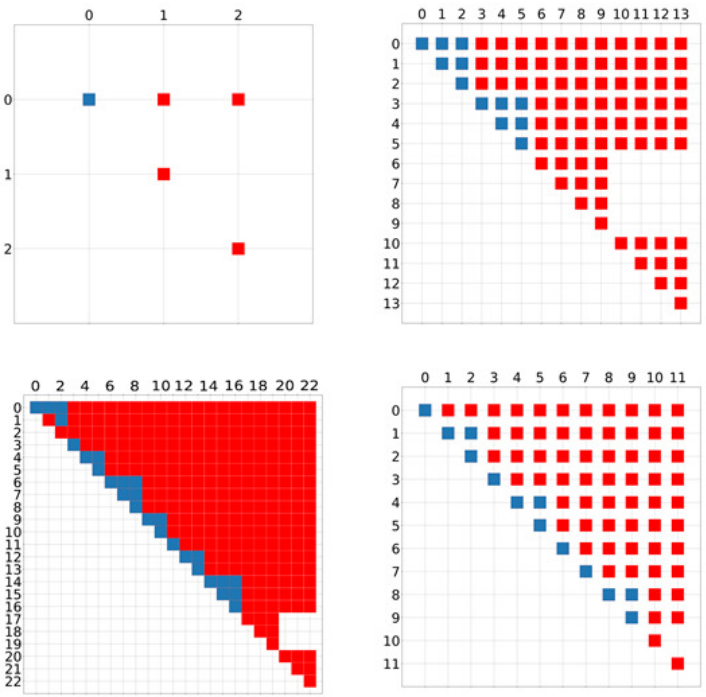


Fig. 2: Overview of QUBO matrices for experiments 1 (upper left), 7 (upper right), 27 (lower left) and 28 (lower right).

Tab. 1: Details of selected benchmark experiments

exp.	$\vec{j}_{max}$	$\vec{pv}$	e_{min}	e_{max}	nr. qubits
1	$[0, 1]$	$[0, 1]$	0	1	3
7	$[7, 5]$	$[10, 3]$	1	10	14
27	$[5, 1, 2, 4, 2, 1, 3, 4]$	$[1, 3, 2, 5, 2, 1, 2, 4]$	2	5	23
28	$[0, 1, 2, 1, 2, 1, 0, 1, 2]$	$[1, 1, 2, 0, 2, 1, 2, 1, 3]$	0	1	12

for the selected experiments 1 to 28. However, the number of qubits increases only from experiment 1 to 27 and it is reduced in experiment 28 compared to 27. This is due the reduction of the maximal number of charging levels j_{max} (from 5 in exp. 27 to 3 in exp. 28). Thus, the largest number of qubits is found for exp. 27 due to a combination of the number of time steps and charging levels.

To study the complexity of the combinatorial optimization program, we show the QUBO matrices of the selected experiments in Fig. 2. The non-zero qubits related to the optimization variable j representing the cost function are marked in blue, whereas the penalties causes the qubits marked in red to be non-zero. Obviously, the most of the non-zero elements in the QUBO matrix originates from the penalties since eqs. (7) and (8) describe global penalties, i. e., corresponds to interactions between all j qubits. Another remarking point is that there are qubits, which are not included in the cost function but represents the slack variables t and v in the penalty terms of eqs. (7) and (8). Their number is determined by the values of the binary representation of parameters e_{min} and e_{max}. Although the cost function includes rather limited coupling of the qubits, it is observable that an increase of the maximal charging levels $j_{max}(t)$ at a certain time step t yield to an interaction of the qubits representing the integer value j as a binary, see the blue marked elements in Fig. 2, e. g., for exp. 7. Since for future applications, a large amount of charging levels is expected, this increase in complexity should be considered for a strategy to achieve quantum advantage with shallow circuits [BGK18] in this special application case.
Since the coupling of the QUBO matrix does not necessarily correspond to a increase in complexity in the optimization landscape for the quantum approximate optimization algorithm [FGG14] with $p = 1$, we also study the landscape in Fig. 3. The optimization landscape corresponds to the expectation value of the sampled cost functions. In Fig. 3, we show the landscapes for experiments 1 (upper part) and 7 (lower part). On the left of fig. 3, the landscape is drawn for the program without penalties, i. e., $C_a = C_b = 0$, and on the right, it is shown considering the penalties. As indicated by the QUBO matrices, the penalties complicate the optimization landscape by increasing the number of local minima but also by increasing the variance of the landscape. Both effects might yield to problems for the classical optimizer of the QAOA algorithm to find the optimal values for the parameters γ and β for the cost Hamiltonian and the mixing Hamiltonian, respectively.

4 Summary and Outlook

We have shown the formulation of a real-world application using MILP and QUBO models and the involved complexity for the QUBO model using a visualization of its matrices and the optimization landscape. The next step will be to analyze if a measure of complexity can be found based on the optimization landscape and if it can be estimated using only the (easily available) QUBO matrix. Since the main part of the coupling terms within the QUBO matrices arises from the global penalties, different implementations of the penalties will be considered in future works. Here, a promising approach seems to be an implementation using conditional gates or a dynamic decoupling approach [De22].
Additionally, we will study if a QAOA algorithm can be used to solve benchmark experiments resembling real-world application scenarios. We will build the algorithm for real quantum hardware and consider different transpiling, mapping and dynamic decoupling strategies to find the most suitable approach for our benchmark. Additionally, we will investigate the

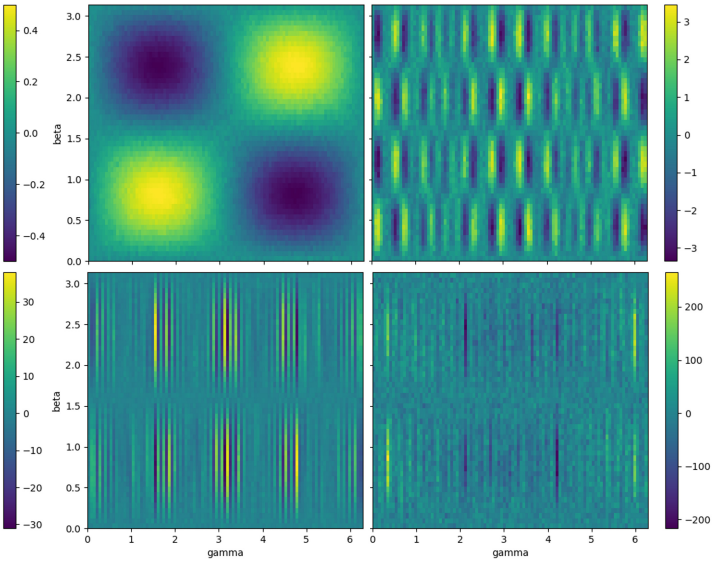


Fig. 3: Energy landscapes for experiments 1 (upper row) and 7 (lower row), without penalties (left column) and with penalties (right column).

influence of the classical and quantum part of the hybrid quantum-classical approach to discover bottlenecks and improvements.

Acknowledgment

This work was financially supported by the Bundesministerium für Wirtschaft und Energie through the project ”EnerQuant” (Project- ID 03EI1025C). We thank Murad Babayev, Andreas Sturm, Fred Jendrzejewski, Yannick Deller, Sebastian Schmitt, Valentin Kasper and Philipp Hauke for fruitful and inspiring discussions.

References

[ABG+20] Abergel, T.; Bunsen, T.; Gorner, M., et al.: Global EV Outlook 2020 - Entering the decade of electric drive? 2020.

[BGK18] Bravyi, S.; Gosset, D.; König, R.: Quantum advantage with shallow circuits. Science 362/, pp. 308–311, 2018.

[De22] Deller, Y.; Schmitt, S.; Lewenstein, M.; Lenk, S.; Federer, M.; Jendrzejewski, F.; Hauke, P.; Kasper, V.: Quantum approximate optimization algorithm for qudit systems with long-range interactions, 2022, URL: https://arxiv.org/abs/2204.00340.

[FGG14] Farhi, E.; Goldstone, J.; Gutmann, S.: A Quantum Approximate Optimization Algorithm, https://arxiv.org/pdf/1411.4028.pdf, 2014.

[Ko14] Kochenberger, G.; Hao, J.-K.; Glover, F.; Lewis, M.; Lü, Z.; Wang, H.; Wang, Y.: The unconstrained binary quadratic programming problem: a survey. J. Comb. Optim. 28/, p. 58, 2014.

[Kü21] Küstner, M.: Modellierung von Problemen als quadratische unrestingierte binäre Optimierungsprobleme, Bachelor Thesis, TU Ilmenau, Department of Mathematics and Natural Sciences, 2021.

[Lu14] Lucas, A.: Ising formulations of many NP problems. Frontiers in Physics 2/5, https://www.frontiersin.org/article/10.3389/fphy.2014.00005, p. 1, 2014.

A new Pattern for Quantum Evolutionary Algorithms

Volker Reers[1], Jörg Lässig[2]

Abstract: Quantum Evolutionary Algorithms have been discussed in the literature in different forms. One branch in these efforts studies approaches for the representation of genetic information, i.e., problem information, in terms of qubits. A typical downside of this representation has been the loss of quantum information in the evaluation and selection steps of the algorithm. I.e., algorithms are implemented in a hybrid-classical setup and require measurements in each iteration. This inevitably destroys superpositions and entanglement structures in the genome representation. In this work, we propose a new implementation approach for genetic information and the evaluation and selection phase, which realizes those steps within the quantum circuit. To achieve this, we utilize qudits for representing the evolving entities. Additionally, we make use of patterns for the design of quantum sub-circuits to compose control structures known from the classical realm. As a result, we show a quantum circuit design for anytime algorithms that does not have to be measured in every iteration and that does not depend on classical control. The overall progress of the evolutionary process only needs to be checked occasionally on a flag-qubit. The approach currently comes with some limitations e.g., in the objective function. It is presented here for the toy problem Leading Ones.

Keywords: Quantum Information Processing, Quantum Evolutionary Algorithm, Quantum Genetic Algorithm

1 Introduction

Evolutionary Algorithms (EA) are stochastic search and optimization methods conceptionally based on natural biological selection and evolution. The three main branches of evolutionary algorithms that have been established in recent years are genetic algorithms (GA), evolutionary programming (EP) and evolutionary strategies (ES).

EAs are based on populations of potential solutions. Particular solutions are also called individuals. EAs apply the principle of survival of the fittest to generate successively better approximations to the optimum. In each generation of the EA, a new population is created by selecting individuals according to their fitness related to a problem to be solved and their reproduction using variation operators. This process leads to the evolution of populations and to individuals that are better adapted to their environment than the individuals from which they originated, i.e., better solutions to the problem to be optimized. This process corresponds exactly to the principle of natural selection and

[1] Qseidon GmbH, Langer Weg 7, Gütersloh, 33332, volker.reers@qseidon.de

[2] Hochschule Zittau/Görlitz, Dept. of Computer Science, Brückenstraße 1, Görlitz, 02826, j.lässig@hszg.de

doi:10.18420/inf2022_98

adaptation to new environmental conditions. Therefore, EAs can be characterized by the representation of the individual, the function to evaluate the fitness of individuals, and the population dynamics, i.e., population size, variation operators, selection of parent individuals, and reproduction and mutation to name a few [HK02].

Quantum algorithms are algorithmic systems based on the laws of quantum mechanics to perform computations. They can be efficiently executed on a quantum computer. In contrast, simulation on a classical computer for the purpose of proof of principle can be very resource intensive, depending on the specific quantum algorithm. It has already been shown that calculating solutions using quantum computers can dramatically improve performance when solving problems such as factorization (see Shor's algorithm) or searching an unstructured database (see Grover's algorithm).

Since Evolutionary Algorithms can basically be described as search algorithms, a quantum version of an Evolutionary Algorithm seems to be a relevant topic for the future when scalable and fault-tolerant quantum computers become available. Moreover, the integration of the two paradigms may be a way to apply quantum computation to difficult problems for which no quantum algorithm is yet available [Ma08].

2 Quantum Evolutionary Algorithms

Evolutionary Algorithms with respect to Quantum Computing have been discussed in the literature since the 1990s. The algorithms that arose over the years can be categorized mainly in three kinds [Zh11]. Evolutionary Designed Quantum Algorithms (EDQA) pursue the goal of automating the synthesis of new quantum algorithms using evolutionary algorithms. Thus, classical evolutionary algorithms are utilized to help developing new or to optimize existing quantum algorithms. As another kind, Quantum Inspired Evolutionary Algorithms (QIEA) concentrate on generating new evolutionary algorithms using some concepts and principles of quantum computing. In a sense, these algorithms are the complement of EDQAs with the goal of creating new evolutionary algorithms for the classical realm. The third kind of algorithms are Quantum Evolutionary Algorithms (QEA), which focus on trying to implement evolutionary algorithms in a quantum computation environment in order to take advantage of quantum computation’s exponential parallelism.

In the following, the focus will be on QEA. QEAs adopt the concepts and principles of quantum computing and therefore use qubits and their entanglement as well as the superposition of states. Like classical evolutionary algorithms, QEAs are characterized by the representation of the individual, the fitness function, and the population dynamics. However, instead of a binary, numerical, or symbolic representation, QEAs use qubits as a probabilistic representation, which is defined as the smallest unit of information in quantum computing. A qubit individual is defined by a set of qubits. The qubit individual has the advantage that it can probabilistically represent a linear superposition of states, meaning binary solutions, in the search space [HK02]. In this sense, a single

qubit individual can represent a population of several binary solutions.

In general, a lot of work in the literature focuses on the representational part, i.e., mapping the genome representation to qubits [Ma08, LC12, Wa13, MT21]. This comprehends for a set of qubits representing a number of individuals probabilistically because such a qubit-based multi-shaped individual represents the superposition of all possible states initially with the same probability. As the multi-shaped individual evolves through the application of quantum gates, it converges from a superposition state to a single state and therefore the variety disappears gradually. This finally leads to a single-shaped individual supposed to represent the optimal solution [HK02].

What all approaches have in common is the representation of information in the terms of the quantum realm. The basic unit of information is therefore the qubit which may be in the basis state $|0\rangle$, $|1\rangle$, or in an arbitrary state between, called superposition. The superposition is for the single qubit case defined by complex values α and β and given by a linear combination of the basis states, described by the equation $|q\rangle = \alpha|0\rangle + \beta|1\rangle$. In this equation, α and β are complex numbers that determine the balance between the two vectors, so that the Born rule is satisfied: $|\alpha|^2 + |\beta|^2 = 1$. Here $|\alpha|^2$ and $|\beta|^2$ provide the probabilities of measuring the qubit in state $|0\rangle$ or $|1\rangle$, respectively. An individual, chromosome or population can then be encoded as a string of qubits of length *n*:

$$\left[\begin{array}{c|c|c|c} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \beta_1 & \beta_2 & & \beta_n \end{array}\right] \tag{1}$$

To modify this entity of information, quantum gates are applied. These conduct rotations on the vector that describes the qubit state [MT21].

A register of qubits for storing information leads to the concept of qudits. A qudit is defined as the quantum version of d-ary digits whose state can be described by a vector in d-dimensional Hilbert space. In this regard, a qudit can be realized by utilizing a set of qubits. The resulting multi-level nature of the qudit allows for a larger state space to store and process information, as well as the ability to perform multiple control operations simultaneously [Wa20]. Although the concept of qudit introduces advantages in various applications and potentials for future development, this system received less attention in the field of QEA and other quantum algorithms than the conventional qubit-based quantum computing. One of the reasons for this is because of technical difficulties in realizing reconfigurable and reprogrammable devices based on qudits rather than qubits [Ku17, Ch22]. As a workaround, a register of qubits can be used to mimic a single qudit.

Considering those registers of qubits as computational entities for algorithms opens a new area of algorithm design including the concept of d-level qudits and control structures, which can be taken advantage of inside the quantum circuit.

3 Patterns for Quantum Control Structures

In classical computing and algorithm design, basic control structures play an important role. They are essential for the control of algorithmic processes. Since these control structures are composed of basic logic gates, the quantum versions of the respective control structures can be composed of quantum gates. With those quantum equivalents of basic control structures, the need for conducting measurements and therefore losing information from quantum states like superpositions and entanglement should no longer be required in each iteration. The following sections describe the control structures needed within the scope of this paper, although this is far from exhausting the field of control structures within quantum computing.

3.1 If-Then Control Structure

An If-Then control structure relies on a condition and an action that is meant to be taken if the condition is fulfilled. In quantum computing, such a structure can be mimicked with a control qubit representing the condition and a qubit on which the action, thus, the application of one or more quantum gates, is applied. Counterintuitively, the action on the respective qubit is directly followed by its inverse, but since the inverse is framed by two CNOT gates controlled by the control qubit, the action will stay resident on the action qubit, depending on the state of the control qubit. Thus, for an If-Then control structure we have to apply the required action and revert it, if the condition is not met, as visualized in Fig. 1. The structure therefore works in the way that an action, in this case a rotation around the Y-axis with angle *θ/2*, is conducted on action qubit q1 and is followed by the inverse rotation around the Y-axis immediately if the control qubit q0 is in state $|0\rangle$. Else, if the control qubit is in state $|1\rangle$, the R_y-gate is followed by an effectively applied NOT-gate which reverses the direction of the Y-rotation. The following inverse therefore rotates in the same direction as the first R_y, adding up to a rotation with angle *θ*. It should be noted that current frameworks for quantum computing like IBM's Qiskit allow for the construction of directly controlled rotations. Since those are based on basic quantum gates, the latter version is illustrated in Fig. 1.

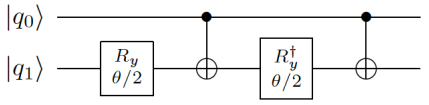


Fig. 1: Controlled rotation mimicking an If-Then Control Structure

3.2 Equals Control Structure

The Equals control structure compares two values and results in an assertion whether the values are equal or not. To have such a control structure in a quantum circuit, the so-called SWAP-Test has been proposed [Bu01]. The SWAP-Test is based on three qubits of which the first two are the qubits to be compared and the third is for the result. The

result qubit consists of a conditional SWAP gate, or Fredkin gate, framed by two Hadamard gates. The conditional SWAP gate in turn can be composed of a Toffoli gate framed by two CNOT gates. Fig. 2 shows the described quantum circuit. Measuring the result qubit will lead to 0 with a probability of 100% when the two compared qubits are equal. In any case where the qubits to be compared are different, the probability of measuring 0 on the result qubit will decrease.

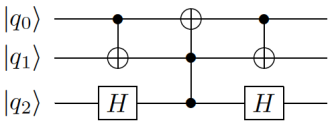


Fig. 2: SWAP-Test mimicking an Equals Control Structure

3.3 Greater-Than Control Structure

To the best of our knowledge there is no possibility to determine whether a single qubit is greater than another qubit in relation to the computational basis without measuring it. This is because the Greater-Than control structure depends on a value for the comparison, which for a single qubit in superposition does not apply. In a sense, every state a qubit can be in is an arbitrary superposition until a relation to the computational basis is applied by measuring it. For this reason, we need more information to realize a Greater-Than control structure. This can be achieved with qudits which act as a multi-level computational unit alternative to the conventional 2-level qubit [Wa20]. A deeper dive into the realization of qudits would be far beyond the scope of this work. To describe the Greater-Than control structure, we assume a qudit as an array of qubits. Since such an array can contain binary values for each qubit, the qudit can be imagined as an integer type. Thus, it contains information that can be compared to another qudit to determine which is the greater one. For the purpose of mimicking a Greater-Than control structure it then is sufficient to calculate the resulting high-bit of the subtraction of the two qudits that have to be compared [Cu08]. In the circuit shown in Fig. 3, the ancilla qubit q_7 represents the high bit. The ancilla is 1 if $Q_1 = [q_0, q_3, q_5]$ is greater-than $Q_2 = [q_1, q_4, q_6]$.

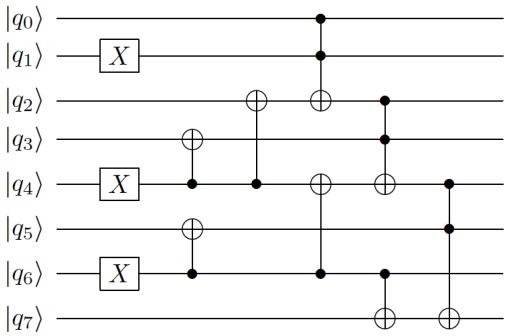


Fig. 3: High-bit computation mimicking a Greater-Than Control Structure

4 Design proposal for Quantum Evolutionary Algorithm

Taking the control structures shown in the previous chapter together, a QEA can be designed as a bigger picture. To be able to utilize all the control structures, we use qudits for our algorithm. We consider three 16-level qudits, each composed of 4 qubits: PQ to represent a state to be approximated, EQ as an evolving population and CQ as an older population to which the newer one shall be compared. We prepare our circuit with a problem representation named PQ, that can contain a precise representation of the problem to be solved or a representation lowered to a threshold, meaning a minimum that has to be achieved by the algorithm. In this regard, PQ can be seen as a termination condition. The evolving population named EQ and the population for comparison named CQ are prepared in superposition state respectively with applying Hadamard gates. The initialization on the left side of the barrier in Fig. 4 is meant to occur only once whereas the gates on the right side of the barrier are meant to be applied in several iterations. As in the classical evolutionary pattern, the next steps consist of generating a new population, evaluating its fitness level and select the one to be taken over to the next iteration. These steps are described in the following subsections.

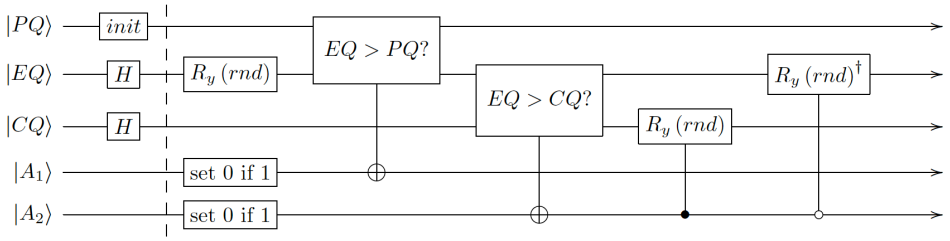


Fig. 4: Design proposal for a fully quantum circuit inherent QEA. The looping component of the circuit is represented on the right-hand side of the barrier.

4.1 Crossover and Mutation

In Evolutionary Algorithms, solutions are created repeatedly, and good solutions are retained to have the algorithm converge against a given target function. This leads to decreasing variety of the population of solutions. To compensate this, Crossover and Mutation are functions to maintain variance in the process of population reproduction. Crossover traditionally creates two new offspring by mixing the genetic information of two parent individuals and allows for larger progresses in the search space as more adjustments are made at once. In this regard, Crossover can be seen as a macro mutation. Mutation on the other hand provides for smaller steps due to more selective changes in the genome representations. Both functions do not compete for better results but rather should be applied in a balanced fashion [BL04].

Our proposal can be defined as a quantum version of (1+1) EA in which the generation of a new population is achieved by creating an arbitrary state on EQ through applying rotations around the Y-axis with random rotation angles θ. The values for θ are created classically and are updated in each iteration to credit for the stochastic character of the algorithm. Whether this relates more to Crossover or to Mutation would exceed the scope of this paper. For our purpose, every application of the rotation gates can be seen as a form of Mutation. Our main focus lies on accommodating classical control structures in the quantum circuit to avoid repetitive measurements at runtime.

4.2 Fitness Evaluation

The evaluation of the Fitness of the population is an important step as it indicates the fitness level of a solution compared to previously generated solutions and with regard to the problem to be solved. In these cases, the Fitness Evaluation can be seen as a representation of evolutionary pressure which sorts out the respective solutions that cannot keep up with other solutions that came up before.

In our proposal, we do not consider invalid solutions but rather focus on evaluating if the respective solution meets the currently reached fitness levels or not. We realize the Fitness Evaluation in our circuit design by comparing the current solution EQ with the termination condition PQ. The latter may be encoded as an exact problem definition or a minimum threshold of the problem that has to be reached by the evolving population. If EQ is greater than PQ, the respective ancilla qubit is set to 1 which can be used as a termination criterion for the algorithm. Otherwise, the ancilla remains 0.

4.3 Selection

To have the Evolutionary Algorithm converge over time, the desired process is to increase the fitness of the solutions in each iteration. This means that for a population $P(t+1)$ a higher fitness level is desired than $P(t)$ showed before [BL04].

We realize this in our approach by simply comparing the evolutionary qudit EQ with the comparison qudit CQ. CQ starts with the same superposition as EQ and serves as a storage qudit in case EQ evolves to better solutions. The comparison in our example is again the Greater-Than control structure, meaning that an EQ in step $t+1$ with a higher value than the CQ from step t will be retained. To achieve this, we utilize the ancilla qubit from the Greater-Than control structure. If the ancilla is 1, indicating that EQ is greater than CQ, then the If-Then control structure from Section 3.1 comes into play. In this case the condition for applying the same gates with the same rotation angles θ is met. Therefore, CQ will be set to the same state that EQ has been set before. This mimics a variable assignment known from classical computing. The same ancilla will be negated right after utilization for keeping the solution of EQ. The negated ancilla is then utilized to decide for another If-Then control structure, namely the inversion of the rotations applied to EQ at the beginning. This conditional inversion is done to maintain a flow of convergence rather than just rolling dice.

5 Results

For the sake of simplicity, we demonstrate our quantum algorithm with the toy problem Leading Ones (LO). LO is defined as a binary string of length n for which the number of *1's* shall be maximised [Su21]. In principle our approach can be applied to other problems with some adaptions. The results are obtained by simulation.

After initialization, a new population will be created by applying a R_y-gate with a random rotation angle θ on EQ. Therefore, EQ is in an arbitrary state. At the same time, the two ancilla qubits are prepared. They will be set to $|0\rangle$ if they equal $|1\rangle$, for which the Equals control structure from Section 3.2 combined with the If-Then control structure from Section 3.1 is utilized. In the next step, the arbitrary EQ will be compared with the problem representation PQ to determine, whether it has reached a desired state. If that is the case, the first ancilla qubit is set to $|1\rangle$, which can be used as a signal to terminate the algorithm if needed. The next step consists of determining whether EQ has reached a better solution than CQ, i.e., if the new population is better than the old one. If that is the case, the new population will be taken over in CQ to retain it for the next iteration. If EQ is not better than CQ, the R_y-gate applied to EQ at the beginning will be inverted, i.e., the new population will be discarded. Now, if no termination takes place, a new iteration follows, and the algorithm starts again with creating a new population for EQ as described above.

To show a proof of principle by means of LO we set the qudit for the problem representation PQ to 1, meaning that all qubits belonging to PQ will be initialized with *1*. Furthermore, we reduce the possible random θ for the variation of every qubit the evolutionary qudit consists of to be either *0* or π and spare the superpositions for EQ and CQ. The task is thereby reduced to reaching a state where all four qubits of EQ are set to one, which by itself is trivial and does not need any quantum device to be solved. But the focus shall be on demonstrating the possibility of shifting all the components needed to

control the flow of the algorithm into a single quantum circuit.

Considering a set of $\theta = \{0, \pi, 0, \pi\}$ for the first operation on EQ, the latter will be greater than CQ, which will be in state $|0\rangle$ after initialization. Therefore, the same operation with the same θ applies on CQ to save the state from EQ. Since the goal of EQ reaching a state of $|1\rangle$ is not yet achieved, the next iteration with new θ will start. The new random set is $\theta = \{\pi, \pi, \pi, \pi\}$, which will lead to EQ in state $|1\rangle$. With this setting, the goal of LO is achieved and the respective ancilla is set to $|1\rangle$. Also, EQ is greater than CQ which is in the state from the previous iteration and will be set to the new state of EQ now. If no termination takes place, the next iteration starts, but since EQ cannot achieve a greater state than CQ anymore, no further changes will occur.

6 Conclusion

We proposed a Quantum Evolutionary Algorithm for which the control of crossover / mutation, fitness evaluation and selection is completely managed within the quantum circuit. Therefore, only minimal interaction with classical components is needed, first and foremost, for the creation of randomized rotation angles to account for the variation of the evolving population and second for the control of the iterations of the quantum circuit itself. As another outcome, we showed how control structures known from the classical realm can be utilized as patterns with which quantum sub-circuits can be constructed. We demonstrated an example flow of the algorithm and pointed out the general mechanics of our circuit design. Our circuit allows for evolving a population mainly in the quantum realm, avoiding measurements during the algorithm's iterations.

Nevertheless, there are some open topics that need further research and development. Firstly, and most important, the algorithm has to be generalized. On the one hand, the algorithm shall be enabled to adapt further variants of EAs in the quantum domain like parallel EA and multi-agent EA and to account for comprehensive runtime analysis [LS14, Ta14]. On the other hand, the goal is to solve real optimization problems. The toy problem Leading Ones suits as a kind of placeholder to just present the proof of principle of the algorithm. But of course, it does not satisfy the requirement for a general optimization algorithm. Further algorithmic development and covering further and practically more important problems will be our main focus for future work. Secondly, the quantum circuit has to be optimized. For now, we just showed that the algorithm works in principle. But the circuit can be optimized regarding the number of used gates and ancilla qubits. Moreover, no quantum error correction was considered in our proposal since we assume noise-free qubits. So, the topic of resources needed is an open research field. And thirdly, the algorithm has been simulated on a classical computer as of today and therefore has to be tested on a real quantum device in our next research steps. This includes the evaluation of several quantum hardware types and tuning the algorithm to the design specifications of the respective hardware.

References

[BL04] Buttelmann, M.; Lohmann, B.: Optimierung mit Genetischen Algorithmen und eine Anwendung zur Modellreduktion. Automatisierungstechnik, Vol. 52, No. 4, pp. 151-163, 2004

[Bu01] Buhrman, H.; Cleve, R.; Watrous, J.; de Wolf, R.: Quantum fingerprinting. Physical review letters, Vol. 87, No. 16, 2001

[Ch22] Chi, Y.; Huang, J.; Zhang, Z.; et al.: A programmable qudit-based quantum processor. Nature Communications, Vol. 13, 2022

[Cu08] Cuccaro, S.; Draper, T.; Kutin, S.; Moulton, D.: A new quantum ripple-carry addition circuit. arXiv preprint quant-ph/0410184, 2004

[HK02] Han, K.; Kim, J.: Quantum-inspired evolutionary algorithm for a class of combinatorial optimization. IEEE Transactions on Evolutionary Computation, Vol. 6, No. 6, pp. 580-593, 2002

[Ku17] Kues, M.; Reimer, C.; Roztocki, P.: On-chip generation of high-dimensional entangled quantum states and their coherent control. Nature, Vol. 546, pp. 622–626, 2017

[LC12] Laboudi, Z.; Chikhi, S.: Comparison of Genetic Algorithm and Quantum Genetic Algorithm. International Arab Journal of Information Technology, Vol. 9, No. 3, 2012

[LS14] Lässig, J; Sudholt, D.: Analysis of speedups in parallel evolutionary algorithms and (1+λ) EAs for combinatorial optimization. Theoretical Computer Science, Vol. 551, pp. 66-83, 2014

[Ma08] Malossini, A.; Blanzieri, E.; Calarco, T.: Quantum Genetic Optimization. IEEE Transactions on Evolutionary Computation, Vol. 12, No. 2, pp. 231-241, 2008

[MT21] Mondal, S.; Tsourdos, A.: Two-Dimensional Quantum Genetic Algorithm: Application to Task Allocation Problem. Sensors, Vol. 21, no. 4: 1251, 2021

[Su21] Sudholt, D.: Analysing the Robustness of Evolutionary Algorithms to Noise: Refined Runtime Bounds and an Example Where Noise is Beneficial. Algorithmica, Vol. 83, pp. 976–1011, 2021

[Ta14] Tao, F.; Laili, Y. J.; Zhang, L.; et al.: QMAEA: A quantum multi-agent evolutionary algorithm for multi-objective combinatorial optimization. Simulation, Vol. 90, No. 2, pp. 182–204, 2014

[Wa13] Wang, H.; Liu, J.; Zhi, J.; et al.: The Improvement of Quantum Genetic Algorithm and Its Application on Function Optimization. Mathematical Problems in Engineering, Vol. 2013, 2013

[Wa20] Wang, Y.; Hu, Z.; Sanders, B.; Sabre, K.: Qudits and High-Dimensional Quantum Computing. Frontiers in Physics, Vol. 8, 2020

[Zh11] Zhang, G.: Quantum-inspired evolutionary algorithms: a survey and empirical study. Journal of Heuristics 17, pp. 303–351, 2011

Influence Estimation In Multi-Step Process Chains Using Quantum Bayesian Networks

Maximilian Selch[1] Daniel Müssig[2] Albrecht Hänel[3] Jörg Lässig[4] Steffen Ihlenfeldt[5]

Abstract: Digital representatives of physical assets and process steps play a decisive role in analysing properties and evaluating the quality of the process. So-called digital twins acquire all relevant planning and process data, which provide the basis, for example, to investigate path accuracies in manufacturing. Each single process step aims to perform an ideal machining after the specification of a target geometry. However, the practical implementation of a step usually shows deviations from the targeted shape. The machine-learning based method of probabilistic Bayesian networks enables the quality estimation of the holistic process chain as well as improvements by targeted considerations of single steps and influence factors. However, the handling of large-scale Bayesian networks requires a high computational effort, whereas the processing with quantum algorithms holds potential improvements in storage and performance. Based on the issue of path accuracy, this paper considers the modelling and influence estimation for a milling operation including experiments on superconducting quantum hardware.

Keywords: manufacturing; path accuracy; digital twin; quantum circuit; quantum algorithm; Bayesian networks

1 Introduction

In recent years, the design if quantum software as well as the construction of real quantum hardware have experienced remarkable progress, which leads to an increasing awareness in a broad range of fields of application to investigate potential operational areas of quantum technologies. The ambitious roadmaps for the upcoming years of major players in quantum computing like IBM [GFW21] and IonQ [Ch20] underline the current dynamics in the enhancements of quantum technologies.

One of these scopes of application is manufacturing, for which quantum computers may help to improve simulations to design and test products, to analyse and work with new materials

[1] Fraunhofer Institute for Machine Tools and Forming Technology IWU, Department Digital Production Twin, Nöthnitzer Straße 44, 01187 Dresden, Germany maximilian.selch@iwu.fraunhofer.de
[2] Fraunhofer IOSB, branch Advanced System Technologies IOSB-AST, Cognitive Energy Systems, Wilhelmsplatz 11, 02826 Görlitz, Germany daniel.muessig@iosb-ast.fraunhofer.de
[3] Fraunhofer Institute for Machine Tools and Forming Technology IWU, Department Digital Production Twin, Nöthnitzer Straße 44, 01187 Dresden, Germany albrecht.haenel@iwu.fraunhofer.de
[4] Fraunhofer IOSB, branch Advanced System Technologies IOSB-AST, Cognitive Energy Systems, Wilhelmsplatz 11, 02826 Görlitz, Germany joerg.laessig@iosb-ast.fraunhofer.de
[5] Fraunhofer Institute for Machine Tools and Forming Technology IWU, Department Digital Production Twin, Nöthnitzer Straße 44, 01187 Dresden, Germany steffen.ihlenfeldt@iwu.fraunhofer.de

doi:10.18420/inf2022_99

as well as to proceed risk and quality modeling in supply and process chains. One issue of the latter topic is to analyse the accuracy of machining processes, to identify dependency relations between the individual characteristics and to search for the decisive influencing factors, which case deviations of the work piece from a preset target geometry. These considerations enable an estimation of the holistic quality of the holistic process chain as well as improvements by considering single steps in a targeted manner. Probabilistic Bayesian networks have proved as an appropriate methodology to reproduce such characteristics of a system. However, with growing network size and complexity, the time required for computations of the system increases exponentially. Especially for handling large-scale networks, quantum algorithms offer a way of a compact representation and potential computational advantages.

After providing an overview of some approaches for quality modeling and influence detection in manufacturing as well as related work in the context of Quantum Bayesian networks, we briefly illustrate the methodology using conventional bit-based computer architecture and how to transfer it to a quantum algorithm. Thereby, we supplement the modeling of Quantum Bayesian networks by an estimator for strengths of influences with the help of SWAP-tests. Afterwards, the presented methods are implemented on superconducting quantum computers based on an exemplary machining application of a milling operation. We evaluate this scenario for a small and larger scale setup and finally discuss on the results obtained from current quantum hardware.

2 Related work

Various statistical and ML-based approaches have been proposed to proceed influence evaluation, deviation analysis and quality improvement of machining processes. For these purposes, methods like regression and classification models [Bu21], fault tree analysis [YBA18] as well as design of experiment techniques [DDT20] are mainly used. These approaches often require some advanced adjustments of the underlying physical process (e.g. empirical formulas for regression). In contrast, Bayesian networks allow to make such statements without previously describing the physics in-depth.

The construction and calculations of Bayesian networks as quantum computing algorithms bases on a representation technique of conditional probabilities which is referred to as qsample encoding [SP18]. The setup of quantum circuits and complexity analysis of qsample encoding for Bayesian networks are described in [ORR13, YC14].

3 Modeling and influence estimation using Bayesian networks

Bayesian networks are directed acyclic graphs, in which nodes model certain influence factors as random variables. If nodes are dependent on each other, they are connected

by an edge, whose direction characterises the dependency relation (pointing from the influencing factor to the dependent variable). If a node is not dependent on any other factor, the corresponding variable holds a probability distribution indicating the likelihood of different status the variable can accept. For this purpose, both discrete as well as continuous variables can be modeled. The dependency relations between the nodes are represented by conditional probabilities. In case of discrete variables these dependencies are stored in conditional probability tables (CPT). When using continuous values, the nodes are usually assumed as normal random variables. Dependencies are modelled by Gaussian distributions, which contain parent-related parameters [Jo01].

The considerations in this article will focus on Bayesian networks with discrete variables for two reasons. Firstly, the influence analysis investigates the impacts of process factors on the basis of pre-defined status, which requires a discretisation of continuous values at some point of the modelling procedure. Secondly, discrete Bayesian networks are particularly suitable for the transmission to quantum algorithmic. In the target application of influence estimations in multi-step machining processes, Figure 1 represents a network example with different part programs of a process chain. In this case, each of the variables is assumed to be binary with probability of $\mathcal{P}_+$ to hold a predefined tolerance criterion for the manufacturing quality resp. a probability of $\mathcal{P}_-$ to not fulfil the criterion. Exemplary, a deviation from a target position with respect to a certain axis can be considered with a threshold value respresenting the quality criterion.

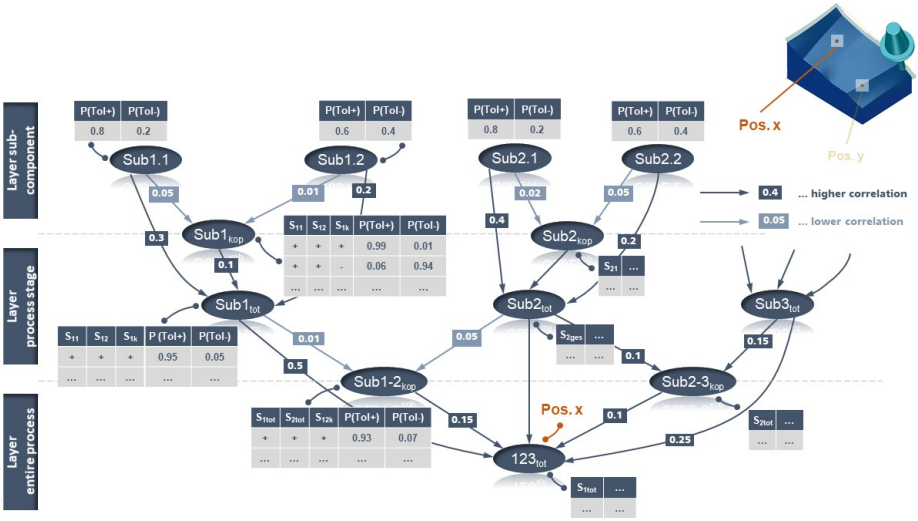


Fig. 1: Example of a Bayesian Network with binary nodes

The structure of a Bayesian network can be obtained by expert knowledge or learned using training data [Jo01]. A Bayesian network with nodes $\mathcal{X} = \{X_1, \ldots, X_m\}$ reflects a unique

joint probability distribution $\mathcal{P}(\mathcal{X})$ which is – in the case of using discrete nodes – given by the product of all conditional probability tables

$$\mathcal{P}(\mathcal{X}) = \prod_{i=1}^{m} \mathcal{P}(X_i | \operatorname{Pa}(X_i)) . \tag{1}$$

In equation (1), $\operatorname{Pa}(X_i)$ describes the set of parent nodes of a variable X_i, i.e. the set of variables on which X_i is dependent on. The single conditional probabilities between dependent nodes are represented as the parameters of the Bayesian network. A parameter

$$\theta_{ijk} = \mathcal{P}(X_i = k | \operatorname{Pa}(X_i) = j) \qquad 1 \leq k \leq s_i, 1 \leq j \leq p_i$$

reflects the probability of a child node X_i to assume if the parent nodes show the status j with s_i depicting the number of different states of X_i and p_i specifying the number of status combinations of $\operatorname{Pa}(X_i)$.

Using classical computing, the parameters of a Bayesian network are commonly learned using the expectation maximization algorithm [PW09]. Given a dataset with $n \in \mathbb{N}$ samples $\mathcal{D} = \{D_1, \ldots, D_n\}$, the goal of the parameter learning is to find the set of parameters $\theta = \{\theta_{ijk}\}$, such that a sample taken from the network matches the data $\mathcal{D}$ best. The Bayesian network can be used to estimate the strength of influences in its dependency relations by analysing the changes in the probability distribution $\mathcal{P}(X_i)$ of a child node X_i when setting evidence to their parent nodes status. The strength of influence of a parent node X_{par} on a child X_{ch} can be described as

$$\mathcal{I}_{X_{\text{par}}}(X_{\text{ch}}) = \frac{1}{\sqrt{2}} \sum_{j=1}^{s_{\text{par}}} \mathcal{P}(X_{\text{par}} = j) \sqrt{\sum_{l=1}^{s_{\text{ch}}} \left(\left[\mathcal{P}\left(X_{\text{ch}} | X_{\text{par}} = j\right) \right]_l - \left[\mathcal{P}(X_{\text{ch}}) \right]_l \right)^2} \tag{2}$$

where s_{par} is defined as the number of different status of X_{par} and s_{ch} as the number of status of X_{ch}. Equation (2) shows the weighted sums of Eucleadian distances of $\mathcal{P}(X_{\text{ch}})$ and $\mathcal{P}\left(X_{\text{ch}} | X_{\text{par}}\right)$ of the different status configurations of the parent node X_{par} – normalised to $\mathcal{I}_{X_{\text{par}}}(X_{\text{ch}}) \in [0, 1]$.

4 Processing Bayesian networks with quantum computing

The calculation of the strength of influence values requires probabilistic inference, i.e. the calculation of posterior distribution of variables when evidence is given by varying the states of parent nodes. This demands a high computational effort in Bayesian networks including a large number of nodes. In the light of the above, this paper focusses on the transfer of this set-up to quantum computing. In a Quantum Bayesian network, each nodes is represented by one or multiple qubits, at which a modelling of s different status of a node requires $\lceil \log_2(s) \rceil$ qubits. Using the technique of qsample encoding, the amplitude vector of a s-qubit quantum state can be used to represent a classical discrete probability distribution. For a probability

distribution of a node X with s different status $\mathcal{P}(X) = p_1, \ldots, p_s,\ (s = 2^n, n \in \mathbb{N})$, measuring the quantum state

$$|\psi\rangle = \sum_{l=1}^{\log_2(s)} \sqrt{p_l}\,|l\rangle \tag{3}$$

in the computational basis is equivalent to sample from $\mathcal{P}(X)$ [SP18]. The encoding of probability distributions for $s = 2$ is done via rotational gates

$$R_Y(\Theta) = \begin{pmatrix} \cos\left(\frac{\Theta}{2}\right) & -\sin\left(\frac{\Theta}{2}\right) \\ \sin\left(\frac{\Theta}{2}\right) & \cos\left(\frac{\Theta}{2}\right) \end{pmatrix}$$

with $\Theta = 2\arcsin(\sqrt{p_2}) =: P_\Theta(p_2)$ to describe $\mathcal{P}(X) = p_1, p_2$ by the probabilities of measuring the state $|0\rangle$ or $|1\rangle$, respectively, of the quantum state $R_Y(\Theta)\,|0\rangle$. Furthermore, to model distributions for $s > 2$, a combination of rotational and controlled rotational is needed to interconnect the required number of qubits to a quantum state representing the node's probability distribution. For example, Figure 2 depicts the circuit to encode $s = 3$ discrete status using the angles

$$\Theta_1 = P_\Theta\left(P\left(|10\rangle\right)\right),\ \Theta_2 = 0, \Theta_3 = P_\Theta\left(\frac{P\left(|01\rangle\right)}{P\left(|00\rangle\right) + P\left(|01\rangle\right)}\right).$$

Thereby, $P\left(|00\rangle\right)$ matches p_1 in $\mathcal{P}(X) = p_1, p_2, p_3$ and $P\left(|01\rangle\right)$ represents p_2 as well as $P\left(|10\rangle\right)$ describes p_3, respectively.

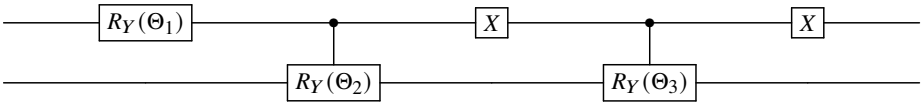


Fig. 2: Quantum circuit to encode a discrete probability distribution with three status

The conditional probability tables of the dependent nodes of the Bayesian network are assembled by a sequence of controlled rotational gates. The controls are the qubits representing the parent nodes and the target is on the child node. Detailed examples of how to proceed these dependency relations in Quantum Bayesian networks are described in [Bo21].

Thus, in comparison with the classical counterpart, Quantum Bayesian networks offer a bunch of advantages in context of memory complexity, itemised:

- compact representation of the network,
- the network's parameters are directly encoded in the qubits representing the nodes,
- logarithmic scaling compared to classical implementation.

Having discussed the build-up of Quantum Bayesian network, a measure for quantify strength of influences analog to the classical way (2) is required. For this purpose, the SWAP-test procedure is suited. The SWAP-test states the certainty of two quantum states to be different from each other. This statement can be transferred to a measure describing the distance between both quantum states. Thereby, a distance of zero corresponds to a probability of 1 of both quantum states to be equal. This probability is given by measurements of ancillary quantum state of the SWAP-test. The other way around, the probability of both quantum states to be different is obtained measuring the ancillary

$$P(\text{anc} = |1\rangle) = \frac{1}{2} - \frac{1}{2} |\langle\psi|\phi\rangle|^2 \tag{4}$$

[Sa20]. Classically, the Euclidean distance between two vectors $\vec{x}$ and $\vec{y}$ representing quantum states calculates as

$$\text{dist}(\vec{x}, \vec{y}) = |\vec{x}, \vec{y}| = \sqrt{(\vec{x} - \vec{y})^2} = \sqrt{|\vec{x}|^2 + |\vec{y}|^2 - 2\langle\vec{x}, \vec{y}\rangle} = \sqrt{2 - 2\langle\vec{x}, \vec{y}\rangle} \tag{5}$$

since vectors representing quantum states are normalised vectors. Comparing both measures, the Euclidean distance (5) correlates positively with the probability (4) of both quantum states to differ from each other. Therefore, to estimate strength of influences in Quantum Bayesian networks, SWAP-tests are proceeded to determine the difference between a specific node with and without giving evidence on parent nodes.

5 Transfer of Bayesian networks to production engineering

In this section, the concept of Quantum Bayesian network is applied to a 3-axis milling operation as a data-driven approach of analysing influence relations between the process factors. The example provides the manufacturing of an aerospace component. The milling operation consists of four processing steps: a face milling, a drilling, a contour milling and end milling operation. The process is visualised in Figure 3 which also shows the assembled number of data records for each single step.

For the setup of Bayesian networks, the following process variables are used

- the axis-specific jerks j_x, j_y, j_z,
- the axial and total position deviations $\Delta_x, \Delta_y, \Delta_z, \Delta_{\text{tot}}$,
- the axis-specific drive currents I_x, I_y, I_z,
- the spindle load L_S,
- the cutting forces according to the Kienzle cutting force model [KV57].

Figure 4 depicts the structure of the corresponding Bayesian network which indicates the dependency relations between the process variables. In [Se21], the results of the strength

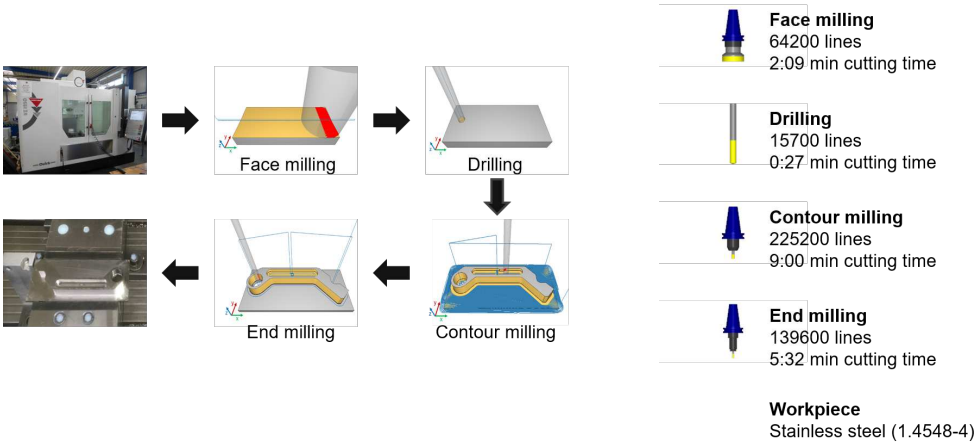


Fig. 3: Manufacturing steps of an aerospace component

of influence estimation via classical Bayesian networks is described. For the evaluation of the quantum computing approach, experiments on the IBM quantum systems are executed using quantum simulators as well as real superconducting quantum hardware devices.

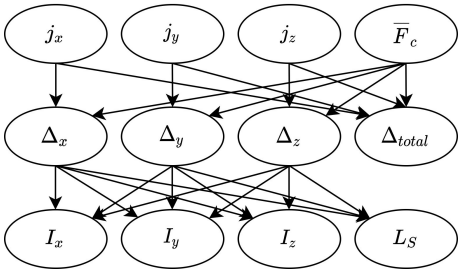


Fig. 4: Structure of the Bayesian network representing the milling application

Quantum Bayesian networks using $s = 2$ and $s = 3$ discrete status for each node were built up. In the $s = 3$ case, only a sub-network of the structure in Figure 4 was implemented in the quantum circuit to be executable at least on IBM's qasm simulator. Figure 5 provides an overview of the calculations for the strength of influence scores to compare the influences of single axial deviations on the spindle load based on (4).

Exemplary, this sub-network is evaluated for the contour milling part which provides the widest variety of machining operations compared to the other sub-processes. In table 1, the strength of influence scores obtained by classical and Quantum Bayesian networks are contrasted. Here, the individual scores can only be compared inside the respective approach, since the positive correlation between the classical and the quantum distance measure does not imply a specific scaling between both scores. Larger values indicate stronger influences

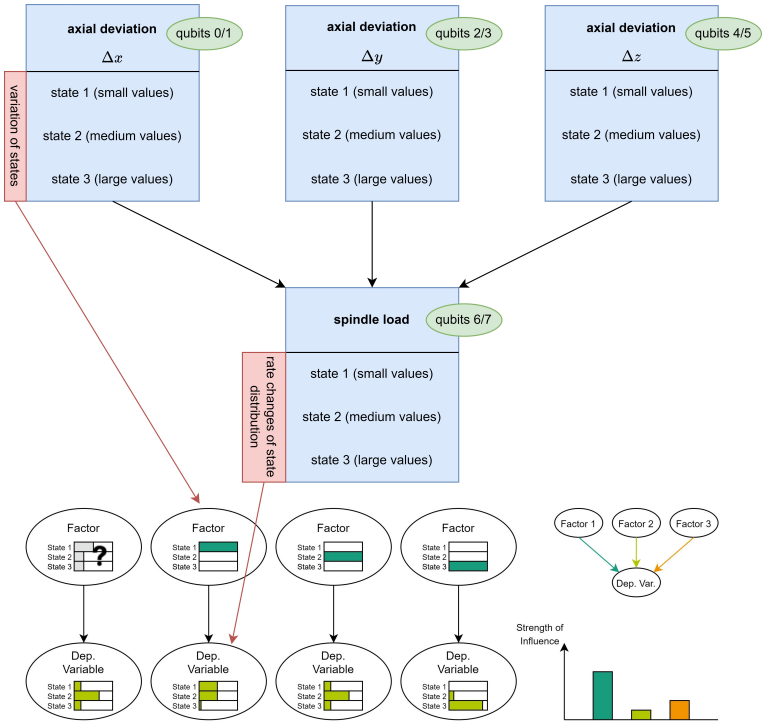


Fig. 5: Setup of the experiments on Quantum Bayesian networks for 3-status nodes

of the respective causal process variables. In both approaches, the influence of the x- and y-axis are dominating the workload. This is in line with the expectations, since these axes correspond to the feed direction of the milling operations.

Approach	Δ_x	Δ_y	Δ_z
Classical Bayesian networks	0.245	0.200	0.161
Quantum Bayesian networks	0.070	0.071	0.059

Tab. 1: Influence scores for process step of contour milling using classical and quantum networks

The execution of the circuits to build up Quantum Bayesian networks and calculating the strengths of influences on today's real quantum computers faces the following device- and circuit-specific problems:

- Real quantum, superconducting devices do not offer an all-to-all connectivity between the physical qubits. To match the circuit with the coupling map of qubits, a transpiling process is required, which increases the depth of the circuit.
- A reduced pool of available quantum gates on the device leads to decompositions of operators, further increasing the circuit complexity.

- The feasibility and accuracy of the circuit depends on the fidelity of the quantum operations (errors of quantum operations).
- The complexity of generated quantum circuits (for example of 3-state-nodes: circuit depth of 1174, width of 40 on IBM qasm_simulator).

6 Using Grover Search

Grover Search was initially framed as an algorithm for searching in unstructured databases [Gr96]. However, at its core, it is a special case of an amplitude amplification algorithm, which can be used in Bayesian networks for interference. The original Grover Search starts in a uniform superposition and searches for one bit-string. In the following years, the algorithm was generalized to find more than one bit-string and to start in any given distribution. Gilliam et al. have shown an extension to the algorithm, where we don't even need to know the searched bit-string. Their algorithm needs to know only if it is the smallest or highest value [GWG20].
For the interference step of the Quantum Bayesian Network (QBN) we start with the probability distribution given by the QBN. Afterward, we apply several Grover-Iterations. A Grover-Iteration consists of an oracle and a diffusion operator. The oracle is a unitary diagonal matrix, which multiplies all bit-strings, which we are looking for, with a -1 and everything else with 1. For example, if we are looking for all bit-strings, where the second qubit is 0 in a 3-qubit system, we need to multiply the following bit-strings by -1: `'000'`, `'001'`, `'100'`, `'101'`.
The second step in a Grover iteration is the diffusion operator, which resembles a reflection of all the amplitudes at their mean value. The operator consists of three steps. The first one is the inverse of creating our initial probability distribution, namely the inverse of our QBN. The third step is to execute the QBN in its natural form. And in between, we need to multiply all amplitudes by -1 except the $|0\rangle_n$ state. The complete diffusion operator is illustrated in Fig. 6. After we have all of our tools together we can execute our interference

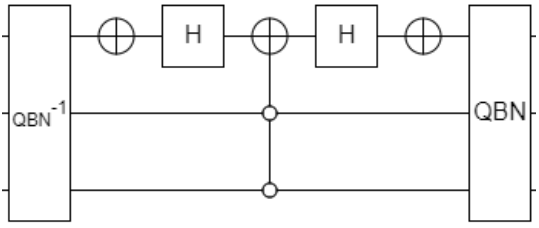


Fig. 6: Diffusor-Operator for a smaller example.

circuit in a hybrid manner. Meaning, that we start with a certain amount of iterations and increase them as we go. We stop, when all of our undesired states are in the region of 0% probability, as Fig. 7 shows. When we compare this approach to the inteference step using a SWAP-Test in Sec. 4 we have a trade off. The implementation with Grover has a high depth,

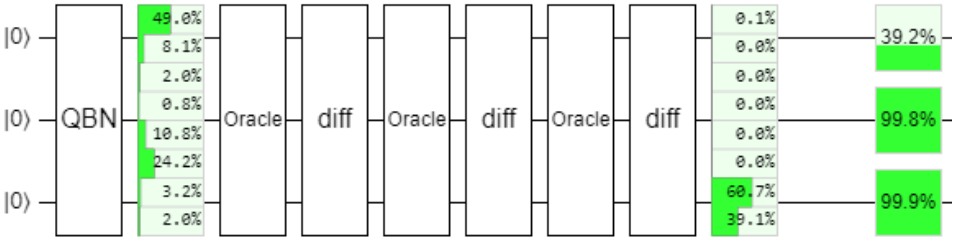


Fig. 7: Grover Serach example for a smaller QBN. The figure shows first the Quantum Bayesian Network as an gate. Afterwards the tool shows the probabilities of all possible states from $|000\rangle$ to $|111\rangle$. Then we apply in this example three Grover iterations, each consisting of the oracle and the diffusion operator. After the iterations we see the new probabilities of all possible states, which represents the queried result.

but does not require more qubits than the Quantum Bayesian Network itself. The inteference step using a SWAP-test, however requires double the qubits (+1) and adds only three gates in depth. Today, neither of both techniques could be executed on an existing quantum device for the given problem. Further, it is expected, that the depth of Grovers algorithm requires a fully fault-tolerant quantum computer. However, in the future, where we have fault-tolerant quantum computers the approach with Grovers algorithm can simulate Bayesian networks with double the size of the SWAP-test approach.

7 Conclusion and outlook

This article focuses on the setup of Quantum Bayesian networks and their application to production engineering, were such networks can be used as a data-driven approach to analyse dependency relations between factors and to estimate the corresponding strength of influences. Quantum computing allows a compact network representation. The discussed potential runtime advantages cannot be proven using the current IBM quantum systems. Even for smaller examples, like the QBN of Fig. 7 without the Grover Search, the circuit depth exceeds the capabilities of current generation quantum computers. The main reason is the multiple controlled *RY*-Gate, which needs to be decomposed into the basis gates of the given quantum computer. Current and future research goals are to implement Quantum Bayesian networks in a more efficient way, analyse further machining operations and to test the developed circuits on a broader range of quantum hardware.

Bibliography

[Bo21] Borujeni, S.E.; Nannapaneni, S.; Nguyen, N.H.; Behrman, E.C.; Steck, J.E.: Quantum circuit representation of Bayesian networks. arXiv:2004.14803, 2021.

[Bu21] Bustillo, A.; Pimenov, D.Y.; Mia, M.; Kapłonek, W.: Machine-learning for automatic prediction of flatness deviation considering the wear of the face mill teeth. Journal of Intelligent Manufacturing, 32:895–912, 2021.

[Ch20] Chapman, P.: Scaling IonQ's Quantum Computers: The Roadmap. IONQ News, 2020.

[DDT20] Deresse, N.C.; Deshpande, V.; Taifa, I.W.R.: Experimental investigation of the effects of process parameters on material removal rate using Taguchi method in external cylindrical grinding operation. Engineering Science and Technology, 23:405–420, 2020.

[GFW21] Gambetta, J.; Faro, I.; Wehden, K.: IBM's roadmap for building an open quantum software ecosystem. IBM Research Blog, 2021.

[Gr96] Grover, Lov K.: A Fast Quantum Mechanical Algorithm for Database Search. arXiv:quant-ph/9605043, November 1996.

[GWG20] Gilliam, Austin; Woerner, Stefan; Gonciulea, Constantin: Grover Adaptive Search for Constrained Polynomial Binary Optimization. arXiv:1912.04088 [quant-ph], August 2020.

[Jo01] Jordan, M.; Kleinberg, J.; Schölkopf, B.; Jensen, F.V.; Nielsen, T.D.: Bayesian Networks and Decision Graphs. Springer New York, New York, NY, 2001.

[KV57] Kienzle, O.; Victor, H.: Spezifische Schnittkräfte bei der Metallbearbeitung. Werkstofftechnik und Maschinenbau, 47:224–225, 1957.

[ORR13] Ozols, M.; Roetteler, M.; Roland, J.: Quantum rejection sampling. ACM Transactions on Computation Theory, 5:1–33, 2013.

[PW09] Pernkopf, F.; Wohlmayr, M.: On Discriminative Parameter Learning of Bayesian Network Classifiers. In: Buntine, W., Grobelnik, M., Mladenić, D., Shawe-Taylor, J. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD, pp. 212–237, 2009.

[Sa20] Sarma, A.; Chatterjee, R.; Gili, K.; Yu, T.: Quantum unsupervised and supervised learning on superconducting processors. Quantum Information & Computation, 23:541–552, 2020.

[Se21] Selch, M.; Hänel, A.; Frieß, U.; Ihlenfeldt, S.: Quality Monitoring Of Coupled Digital Twins For Multistep Process Chains Using Bayesian Networks. Proceedings of the Conference on Production Systems and Logistics : CPSL 2021, pp. 415–425, 2021.

[SP18] Schuld, M.; Petruccione, F.: Supervised Learning with Quantum Computers. Quantum Science and Technology, 2018.

[YBA18] Yahmadi, R.; Brik, K.; Ammar, F.: Causal tree analysis for quality control of the lead acid battery manufacturing process. International Journal of Energy Research, 42:1738–1759, 2018.

[YC14] Yoder, T.L.and Low, G.H.; Chuang, I.L.: Quantum inference on Bayesian networks. Physical Review A, 89(062315), 2014.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1175

Quantum Computer and Quantum Algorithm Benchmarking

Daniel Müssig[1], Jörg Lässig[2]

Abstract: Quantum computing is an exciting field, which brought out many competitors, both on the hardware and algorithmic side. Although users are not expecting to outperform classical machines yet, they are still interested in quantum computers, which can demonstrate advantages of their algorithms best [Ar19]. However, for users, especially in the industry, it is not easy to find a quantum device, which fits their needs. This problem arises from an unmanageable amount of meta information about qubits and connections between them such as fidelity, error rates and timings. In this paper we propose a web tool, which helps the user to find quantum computers best suited for their problem or quantum algorithm. Further, it will generate necessary graphs and reports for benchmark studies, which can be directly included in papers of scientists.

Keywords: Quantum Computing; Quantum Device; Quantum Algorithm; Benchmark Study

1 Introduction

While we are still in the early stage of quantum computing and current systems are still far away from practical relevance, concerning their performance, a lot of research is done to make use of them as early as possible. However, with the sheer amount of available systems utilizing different technologies and architectures, it is hard to choose the right system for the job at hand. Similarly, the performance of near-term quantum algorithms depends heavily on the quantum device. Therefore it is important to choose the right system complementing the algorithm. Although studies that compare a single algorithm on different systems exist, they are quite rare. Thus, a single point of contact for comparative benchmarks could help researchers and even practitioners from industry. Further, this software system that we dub QBench, should be able to conduct benchmarks over all available systems and create reports in different output formats. The use cases and the concept of this system will be discussed in this paper.
The remainder of this paper is organized as follows: In Sec. 2 we discuss existing benchmarks in quantum computing, which influenced QBench. Use cases of the system are presented in Sec. 3. In Sec. 4 the system architecture is discussed. We will conclude in Sec. 5.

[1] Fraunhofer IOSB, branch Advanced System Technologies IOSB-AST, Cognitive Energy Systems, Wilhelmsplatz 11, 02826 Görlitz, Germany daniel.muessig@iosb-ast.fraunhofer.de
[2] University of Applied Sciences Zittau/Görlitz, Brückenstraße 1, 02826 Görlitz, Germany jlaessig@hszg.de

doi:10.18420/inf2022_100

2 Related Work

One of the first metrics, namely Quantum Volume, for comparing quantum devices was created by IBM [Cr19] and was rapidly adopted by several hardware providers and used in news articles. The benchmark behind this generalizes current hybrid quantum circuits and tries to give an answer to the question which circuits can be executed on a given device. It tests therefore the deepest and widest circuits that perform above a certain threshold and reports the lowest of both numbers. However, while a single-number metric is easier to compare by end-users, Quantum Volume lacks a deeper understanding of the hardware, since most of the real-world circuits are not square. Further, the usage of artificial circuits resulted in optimizations on the software side to reach higher scores [Ju20].
An extension of this is volumetric benchmarking, which was recently introduced [Lu21]. The authors focused more on application benchmarking rather than generating artificial circuits and thus giving a first comparison with their benchmarks, which technology has its strengths on which algorithms. While their volumetric benchmarking diagrams are excellent for an overview they are still only reproducible, when the authors of the benchmarks release their code and calibration data of the system with it.
More recently Quetschlich et al. released a web tool, where users can download unified benchmark circuits [QBW22] for, at the moment, two different hardware providers. While it provides unified circuits, there is no mechanism to track changes after downloading it. Further, the results can be only reported by the users and thus there is no way to compare all results between the different systems at once.

3 Use Cases

When thinking about the requirements of a benchmarking tool for quantum computing we came up with the following use cases for different user groups. First, a user has a *problem* that he wants to solve. He is looking for a suitable quantum computer that can best solve the problem with *any algorithm*. In this situation, accuracy plays a major role for him. Another user has a *problem* that he would like to solve in the future on a *quantum computer*. Therefore he wants to know at which size the quantum computer can outperform classical algorithms. Further, a user might want to perform experiments on a *quantum computer*. In this situation, he is looking for a provider that meets his needs as cheaply as possible and has short queues. A scientist who has developed an optimization *algorithm* wants to test whether it performs better than others. To do this, he selects a few *problems* and runs them on *all available quantum computers*. He receives a report from the system with many diagrams and graphs that he can include in his scientific paper. Another scientist has developed a *new quantum computer* that is supposed to work particularly well for wide quantum circuits and now wants to test, whether it is better than the competition. To do this, he selects suitable *problems* and *algorithms* and runs them on his QC. And lastly, a user might want to test a new *algorithm* for error correction and is looking for a *system* with a suitable architecture and sufficient qubits. Lastly, a user might want to conduct benchmarks, where neither the

source code nor the results are shared and/or the *system* requires a user-specific token.
We have decided to benchmark problems and algorithms on quantum devices separately because we think especially the industry is interested to buy access to the right quantum computer for their specific problems. Further, we can provide standardized benchmark problems to compare new and existing algorithms.
Another issue we want to tackle with our tool is the comparison of quantum devices of a single or multiple providers. The ultimate goal is to create a single point, where everyone can compare architectures and important device-specific metrics. Therefore a user will be able to find a device with a fitting architecture, number of qubits, expected executable depth, or fast access over multiple providers on a single site.
On the other hand, we want to help researchers create comparable benchmark studies. First, a researcher shall be able to compare his algorithm on a couple of problems over different quantum devices and receives a report, which he can easily include in his scientific paper. Second, he can use the benchmark platform to share his implementation and interactive comparable graphs with his paper. Further, the benchmark tool will also capture calibration data at the time of the benchmark, to verify the results on a classical device.
Lastly, we envision a verified ranking of quantum devices and algorithms. This is a standard procedure when comparing CPUs, but the quantum sector could also profit from it. In a verified ranking only freely available frameworks will be used. The code will be made public and classical optimization techniques like measurement mitigation or layout optimization as well as, e.g. dynamical decoupling will be tagged.

4 Concept

To implement the features we described in Sec. 3, we envision a web tool, which is deployed in a Kubernetes cluster. We decided to create a web tool to make it as public and easy to use as possible. Thus, the proposed tool does not have any requirements on the hardware of the user. And as discussed before, code can be made publicly available as well as interactive result graphs or verified rankings. The complete system is shown in Fig. 1. We propose to implement the web tool in a Kubernetes cluster to create individual containers, which execute the source code of the users. This will, on the one hand, enable scaling and on the other hand, separate the source code of the users from the web application to reduce the risk of vulnerabilities. The user shall be able to use the Python quantum computing SDK of his choice, such as qiskt, to execute benchmarks. Thus, we will provide a universal benchmark SDK, which the user is required to add to his code. This SDK also encapsulates problems and algorithms to benchmark. Further, we intend to include as many quantum computers from different providers as possible out of the box. However, other providers and quantum computers can be added by the user.

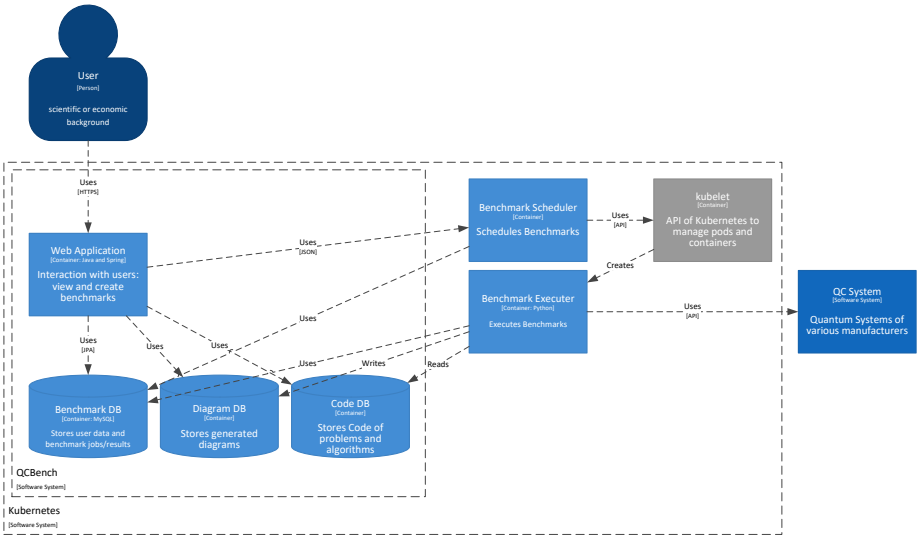


Fig. 1: System overview

5 Outlook and Further Work

We proposed a web tool for benchmarks of problems, algorithms and systems. The tool shall be easy to use and automates as much as possible e.g. generation of scientific reports. The next step is to create a prototype, and if it attracts users to fully implement QBench.

Bibliography

[Ar19] Arute, Frank et al.: Quantum Supremacy Using a Programmable Superconducting Processor. Nature, 574(7779):505–510, October 2019.

[Cr19] Cross, Andrew W.; Bishop, Lev S.; Sheldon, Sarah; Nation, Paul D.; Gambetta, Jay M.: Validating Quantum Computers Using Randomized Model Circuits. Physical Review A, 100(3):032328, 2019.

[Ju20] Jurcevic, Petar et al.: Demonstration of Quantum Volume 64 on a Superconducting Quantum Computing System. arXiv:2008.08571 [quant-ph], 2020.

[Lu21] Lubinski, Thomas; Johri, Sonika; Varosy, Paul; Coleman, Jeremiah; Zhao, Luning; Necaise, Jason; Baldwin, Charles H.; Mayer, Karl; Proctor, Timothy: Application-Oriented Performance Benchmarks for Quantum Computing. arXiv:2110.03137 [quant-ph], December 2021.

[QBW22] Quetschlich, Nils; Burgholzer, Lukas; Wille, Robert: , MQT Bench: Benchmarking Software and Design Automation Tools for Quantum Computing, 2022. MQT Bench is available at `https://www.cda.cit.tum.de/mqtbench/`.

(Agiles) Enterprise Architecture Management in Forschung und Praxis

Infrastructure anomaly detection: A cloud-native architecture at Germany's Federal Employment Agency

Gebhard Herget[1], Eldar Sultanow[2], Alina Chircu[3], Johannes Ludsteck[4], Sebastian Hammer[5], Christian Koch[6], Willy Reuter[7], Matthias Seßler[8]

Abstract: In prior research we explored the use of time series analysis methods to detect one class of information technology (IT) infrastructure anomalies - Distributed Denial of Service (DDoS) attacks. The results of this prior work were a mathematical model and a prototype implementation that were concretely trialed and operated in the data centers of Germany's Federal Employment Agency (FEA). With this paper, we go one step further and generalize as well as optimize the mathematical model and create higher performance and scalability for an updated prototype through targeted use of cloud technologies. The starting point of our generalization is the Exponential Smoothing (E-S) approach, which underlies, for example, the well-known Holt-Winters method. This method is used to predict univariate time series. To detect anomalies (such as DDoS attacks) in infrastructure data, we extend the E-S approach to enable it to forecast multivariate time series. In this optimization of our method and our prototype, we take an exploratory, agile approach. Furthermore, we present a cloud-native architecture stack which we pilot in Azure.

Keywords: Time Series; Enterprise Architecture; Exponential Smoothing; Cloud; Azure; Distributed Denial of Service; Infrastructure Anomaly Detection System

1 Introduction

The availability, integrity and security of an organization's information technology (IT) infrastructure are paramount for providing IT services internally and externally to various organizational stakeholders such as employees or customers. IT infrastructure anomalies, i.e. deviation from normal system behavior that result from outside malicious threats, internal errors, or atypical usage [Fe19], are prime concerns for companies to identify

[1] IT Dept. of Germany's Federal Employment Agency, Tafelhofstraße 4, 90443 Nuremberg, Germany gebhard.herget@arbeitsagentur.de
[2] Capgemini, Bahnhofstraße 30, 90402 Nuremberg, Germany eldar.sultanow@capgemini.com
[3] Bentley University, 175 Forest Street, Waltham, MA 02452-4705, USA achircu@bentley.edu
[4] Institute for Employment Research (IAB), Regensburger Straße 104, 90478 Nürnberg, Germany johannes.ludsteck@iab.de
[5] Capgemini, Bahnhofstraße 30, 90402 Nuremberg, Germany sebastian.hammer@capgemini.com
[6] Technische Hochschule Nürnberg Georg Simon Ohm, Kesslerpl. 12, 90489 Nuremberg, Germany christian.koch@th-nuernberg.de
[7] Microsoft Deutschland GmbH, Unter den Linden 17, 10117 Berlin, Germany willy.reuter@microsoft.com
[8] IT Dept. of Germany's Federal Employment Agency, Tafelhofstraße 4, 90443 Nuremberg, Germany matthias.sessler@arbeitsagentur.de

doi:10.18420/inf2022_101

and address. Distributed Denial of Service (DDoS) attacks are an especially important and widespread type of malicious attack towards a company's IT infrastructure. Internet security reports indicate there were 9.7 million DDoS attacks in 2021 - which represents a slight decrease from 2020 numbers (likely a result of the COVID-19 global pandemic) but is significantly higher than 2019 DDoS numbers [NE22]. DDoS attacks are cheap to launch (sometimes even free), increasingly target specific industries such as voice over IP, computer manufacturing, software publishers, insurance agencies and brokerages, and colleges, universities and professional schools, and generate significant economic losses - sometimes to the tune of tens of millions of dollars - for the affected companies [NE22].

In prior work by Ludsteck et al. [Lu21], we proposed an architecture to detect anomalies in the IT infrastructure and described the implementation of a prototype which was tested in the context of detecting Distributed Denial of Service (DDoS) attacks for the data centers of Germany's Federal Employment Agency (FEA). Since then, the prototype has been in use for about a year, yielding additional insights regarding both technologies and data analysis. In this paper, we build on the lessons learned from the prototype and present an optimized architecture for DDoS anomaly detection at FEA. The contributions of this paper are twofold: the generalization of the time series analysis method used for IT infrastructure anomaly detection and the cloud-enabled scalable architecture. With respect to the first point, we show how the proprietary time series analysis implementation can be replaced with a generalized multivariate exponential smoothing method. With respect to the second point, we discuss a trial of a cognitive cloud service from Azure.

2 Related Work

Our prior work reviews the main characteristics of IT infrastructure anomalies and the related anomaly-based detection systems [Lu21]. Other authors have offered comprehensive surveys of these issues (see, for example, [Fe19]). In this section, we review additional work relevant to the design of the anomaly detection system - namely to the prediction of abnormal IT infrastructure events through time series, which will be used to optimize our prototype.

He et al. [He20] address the limitation of many existing studies on time series prediction, which consider temporal patterns without regard to correlation among variables, thus causing loss of information and inevitably producing false positives. The authors design an unsupervised method to detect anomalies in multivariate time series, which consists of two steps – a prediction step and a threshold selection step. In the first step, they utilize a multi-scale convolution and graph attention network for capturing information in temporal pattern with feature pattern. In the latter step, they identify the threshold value by an MSE (mean square error) based extreme value analysis between the predicted and the observed value.

Li et al. [LPJ17] introduce a framework for anomaly detection in multivariate time series that is based on Hidden Markov Models. Their main idea is to convert multivariate time series into univariate ones. For this, the authors consider conversion algorithms that include fuzzy C-Means (FCM) clustering and fuzzy integral. They apply a Hidden Markov Model (HMM) to discover the presence of anomalies in multivariate time series.

In related work, Li et al. [Li20] present an approach using fuzzy clustering technique to reveal anomalies with respect to the amplitude and shape of multivariate time series. In the first step, the authors use a sliding window in order to construct an array of multivariate sub-sequences. In the second step, they employ an enhanced fuzzy clustering for finding a pattern within these obtained sub-sequences. In the third and final step, they use a reconstruction criterion to recreate the sub-sequences involving the optimal cluster centers and partition matrix. Using a confidence index, they quantify the degree of any anomaly found and tune the detection itself via particle swarm optimization. The authors successfully validate their approach using experiments on various synthetic and genuine data sets.

Jones [Jo66] develops an exponential smoothing method for multivariate time series. His method recursively "estimates the optimum weight matrix for the exponential smoothing and prediction of multivariate time series" and "generalizes to non-linear processes when the non-linear structure is known" [Jo66].

Pfeffermann and Allon [PA89] introduce a "multivariate extension of the familiar exponential smoothing procedure for forecasting univariate time series composed of level, seasonality and irregularity". Using an univariate smoothing procedure together with corrective factors relying upon information from the other series, the authors deduce the up-to-date level, trend, and seasonal effect estimates of each series as weighted estimate averages. Moreover, they use two actual bivariate time series to demonstrate and benchmark the procedure's performance against that of other univariate and multivariate prediction methods.

Zhao et al. [Zh20] address one of the current major limitations of anomaly detection on multivariate time-series: "they do not capture the relationships between different time-series explicitly, resulting in inevitable false alarms". To close this gap, the authors develop a new self-supervised method that recognizes anomalies in multivariate time series by treating each univariate time series as a separate feature and by simultaneously incorporating two graph attention layers designed to learn the (non-trivial) interdependencies between multivariate time-series across both temporal and feature dimensions.

3 Using A Cognitive Time Series Service with Recurrent Neural Network for Infrastructure Anomaly Detection

The work of Zhao et al. [Zh20] forms the basis of Microsoft's Cognitive Time Series service. We used this service to optimize the architecture described in Ludsteck et al. [Lu21]. We

fed our FEA infrastructure data to the service using the Azure Data Explorer[9], which is part of Microsoft Azure Cloud.

For this purpose, first an Azure Data Explorer cluster had to be setup on the Azure Cloud. This allowed the capacity of the system to be expanded dynamically, as new instances could be created on the fly. This is particularly advantageous for later productive usage. Furthermore, the system included a database for both historical data as well as for newly arriving real-time data.

The Azure Data Explorer offers various options for feeding data into the system. One is to stream real-time data via an Event Hub or IoT Hub. In addition, the upload of historical data via file upload is possible. Since we did not have a connection to a system from which data can be streamed yet, we used the file upload. For this, the data was brought into a file format supported by Azure Data Explorer. We chose .csv for this purpose. The converted file could then be uploaded via the Data Explorers web interface.

The Data Explorer offers the option to prepare or plot the data using KQL (Kusto Query Language)[10]. Listing 1 shows the code for extracting anomalies from the data. Thereby, the considered time period was defined with min_t and max_t and the time interval was specified with 10 minutes. The anomalies could then be determined by using the function *series_decompose_anomalies* and then plotted using an anomaly chart. As parameters we used the calculated Mahalanobis distance (a multivariate distance metric) and a anomaly threshold of 7.0.

```
let min_t = datetime(2022-06-19 22:00);
let max_t = datetime(2022-06-20 23:00);
let dt = 10m;
monitoring_data
| make-series num=avg(['mahalanobis']) on Timestamp from min_t to max_t step dt
↪  by Country
| where Country == "*****"
| extend (anomalies, score) = series_decompose_anomalies(num, 7.0, -1, 'linefit')
| render anomalychart with(anomalycolumns=anomalies)
```

Listing 1: KQL (KUSTO Query Language) Query for Azure Data Explorer

An example of anomalies determined by the Azure Data Explorer is shown in Figure 1 with anomalies indicated by red dots. The blue line shows the Mahalanobis distance and the red line the calculated anomaly score. A total of four anomalies were detected when the threshold was set to 7.0. These anomalies correspond in time to the alerts shown in Figure 5.

[9] https://azure.microsoft.com/de-de/services/data-explorer/
[10] https://docs.microsoft.com/de-de/azure/data-explorer/kusto/query/

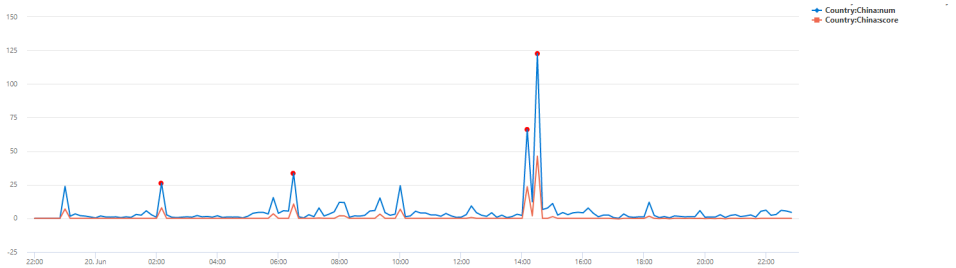


Fig. 1: Anomaly detection graph by Azure Data Explorer (Anomalies are marked with red dots)

Our approach is also capable of running in the Microsoft Azure Cloud. The next steps would be to connect the real-time system to the Azure Data Explorer using the Event Hub and to implement automatic notifications when anomalies are detected.

4 Architectural Findings from the Prototype

The prototype has now been running for almost one year and we gained valuable lessons from its operation.

One of these lessons is that **the Mahalanobis distance is very suitable to detect outliers in the multidimensional time series data of DDoS infrastructure anomalies.** Kotu and Deshpande [KD18] indicate that "[m]ore advanced statistical techniques take multiple dimensions into account and calculate the Mahalanobis distance instead of the standard deviations from the mean in a univariate distribution". Indeed, the "Mahalanobis distance is the multivariate generalization of finding how many standard deviations away a point is from the mean of the multivariate distribution" [KD18]. To detect cyber-attacks on water distribution system, Gjorgiev and Gievska [GG20] employ deep learning architectures based on Variational Autoencoders using Mahalanobis distance as well.

But why is Mahalanobis distance so helpful in our case? The current problem with monitoring is that metrics are surveilled using static thresholds. These thresholds (limit values) must be defined upfront, entered into a monitoring tool accordingly, and then assigned to the metrics/components. The definition of these limit values can be complicated in several ways. In the case of technical threshold values such as CPU temperature or the fill level of a file system, they are usually still easy to determine. For more complex metrics such as the number of accesses to a service or permissible error rates, determining the threshold values can be much more difficult. In addition, threshold values are often dynamic and depend on other metrics.

For this purpose, the Mahalanobis distance method was used as a solution. The metrics used by our prototype were based on aggregates from access logs over a defined time interval of 10 minutes (6 measure points per hour) each.

- *awz*: Response time (mean value)
- *awz.std_dev*: Variance of the response time
- *req*: Number of accesses (requests)
- *ips*: Number of clients (unique IPs)
- *bytes*: Bytes transferred

The method was initially applied to historical data in order to determine its suitability based on existing logs. The analysis was based on data sets of one day in each case. Thus, 5 metrics with $6 \cdot 24$ measurement points were analyzed. The metrics had further attributes such as response codes (200, 404, 500 etc.) or observed services. With 30 services and approx. 30 response codes and further attributes, this resulted in up to: 5 metrics · 30 vhosts · 30 response codes · 200 countries (sources) = 900, 000 time series per day.

For the 5 mentioned metrics the Mahalanobis distance was calculated. This resulted in 30 vhosts · 30 response codes · 200 countries (sources) = 180, 000 attribute combinations. Note that this is the worst case and in practice we handled about 5000 attribute combinations. For each of these 5000 attribute combinations, 5 time series are available.

Using three nested for-loops (over *vhost*, *response_code*, *country*) we went through each attribute combination and considered one time series, for example: all cases where $response\_code = 200$ and the request originated from Italy targeting service "X". And from this time series we picked out the outliers by applying the Mahalanobis distance. For an alert, the statistical significance of each value was determined. The p-value for each distance was calculated as the p-value corresponding to the chi-squared statistic of the Mahalanobis distance with $k - 1$ degrees of freedom, where k is the number of variables. In our case, we used $5 - 1 = 4$ degrees of freedom.

Then warnings were triggered at a Mahalanobis distance of more than 25 and a p-value lower than 0.001. These warnings were summed up on the time axis and if more than 10 alarms were exceeded at the same time, a problem was flagged.

The parameters described above were determined based on the evaluation of different scenarios and are subject to further optimization in the future. The result is a strongly condensed representation of anomalies, which facilitates a visual evaluation via a time series plot. Parameterization is only required at a few points, which allows monitoring of such systems with reasonable effort. Retrospectively, it was possible to detect an anomaly that was not flagged by the currently installed safety mechanisms. A further optimization of our prototype lies in the smoothing of the time series (by using a Kalman filter for example).

In practice, how does the DDoS detection work? Figure 2 shows the result of aggregation over 10 minutes intervals of the utilization data containing the *vhost* and *country*. That

means all requests from all *countries (sources)* on all *vhost* (5000 attribute combinations) are shown in this example given by Figure 2. This specific example is from 2022-06-20 to 2022-06-21.

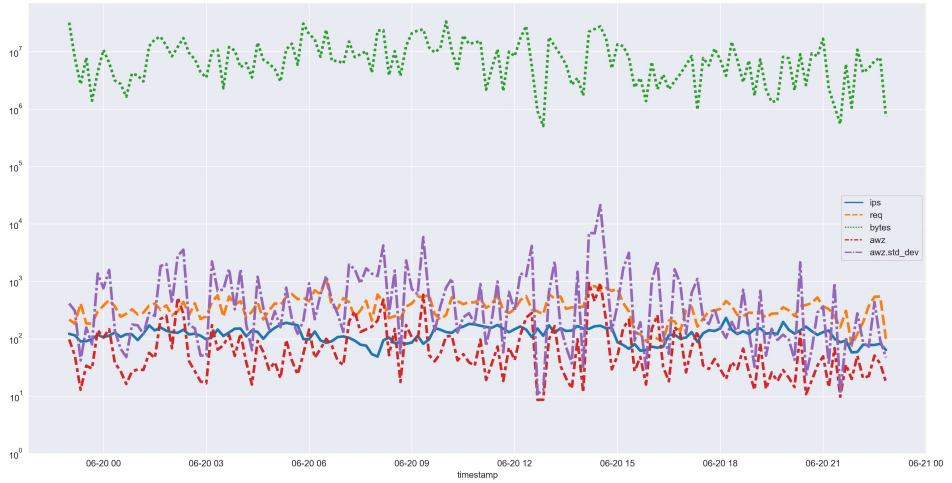


Fig. 2: Aggregation (10 minutes intervals) of of the utilization data containing the *vhost* and *country* (20th to 21st of June 2022)

Figure 3 shows the same raw data as shown in Figure 2. But Figure 3 additionally contains the calculated values of the Mahalanobis distance between the 5 time series (ips number of unique IP addresses, req number of requests, $bytes$ number of bytes transferred, awz response time, $awz.std\_dev$ standard deviation of response time) depicted by Figure 2. This Figure 3 shows the Mahalanobis distance plot as the pink curve, the p curve as the grey curve (the p curve is the Chi-squared test curve - which we obtain by applying Chi square Test on the Mahalanobis distance curve), the $alert$ curve is the result of filtering the p curve together with the Mahalanobis distance curve (the p value must be lower than 0.001 and the Mahalanobis distance value must be larger than 25).

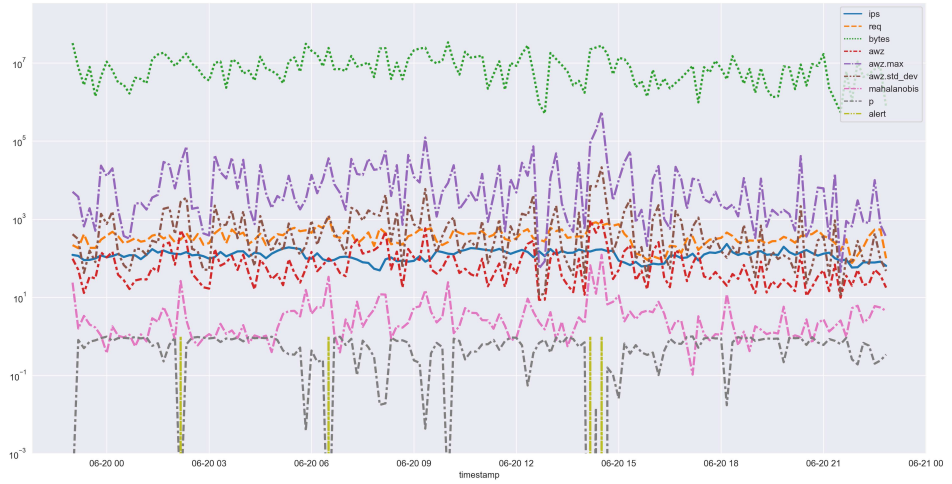


Fig. 3: Example from Figure 2 extended by Mahalanobis Distance curve, p curve and identified alerts.

To get a clearer view, Figure 4 shows only the Mahalanobis distance curve, p curve and identified alerts, which we added to Figure 3.



Fig. 4: only the Mahalanobis distance curve, p curve and identified alerts, which we added to Figure 3 (20th of June 2022)

Figure 5 shows a pair plot (scatterplot matrix) over the five parameter values ips (number of unique IP addresses), req (number of requests/accesses), $bytes$ (number of bytes transferred), awz (response time), $awz.std\_dev$ (standard deviation of response time). The outliers – which are the alerts we identified (by p value being lower than 0.001 and the

Mahalanobis distance larger than 25) – are highlighted in orange. And these orange outliers refer to the (downward) peaks in Figure 3 and Figure 4.

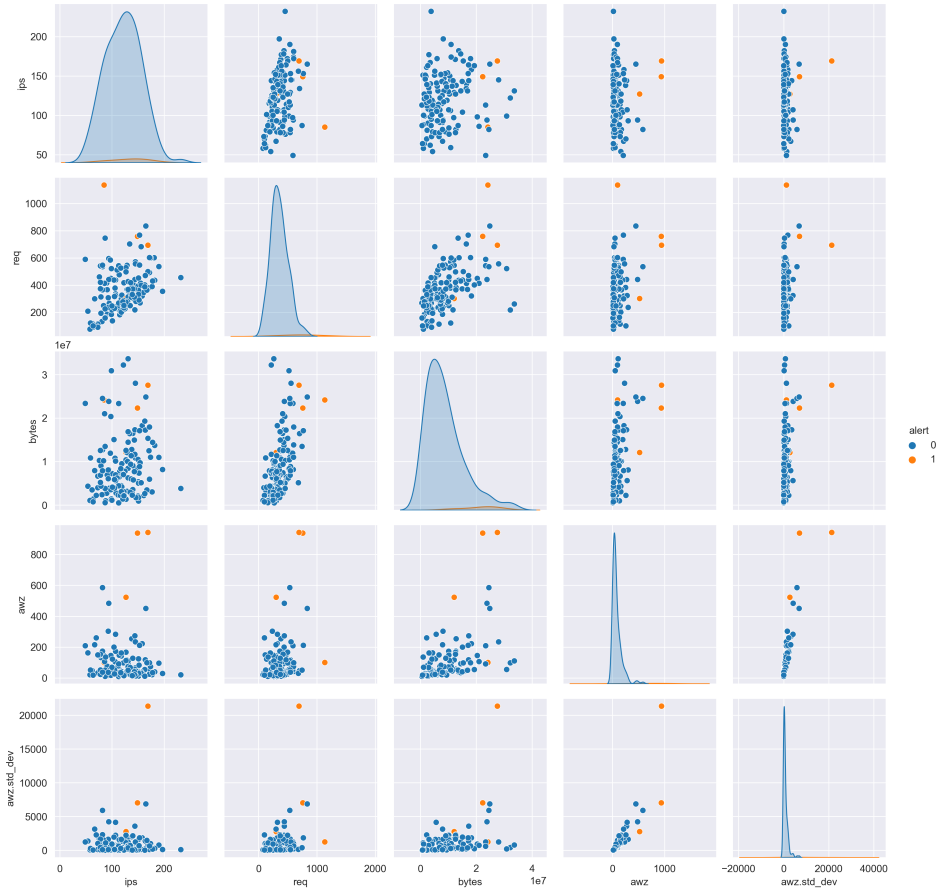


Fig. 5: Pair plot / Scatterplot matrix over the 5 paramter values (ips, req, $bytes$, awz, $awz.std\_dev$)

Figure 6 accumulates all identified alerts in order to determine wether or not there is any prominence in the density of the anomalies within the considered/observed time period (in this case, from 20th to 21st of June 2022). In this plot given by Figure 6 we see two prominent peaks worth taking a closer look at, which we do in Figure 7.

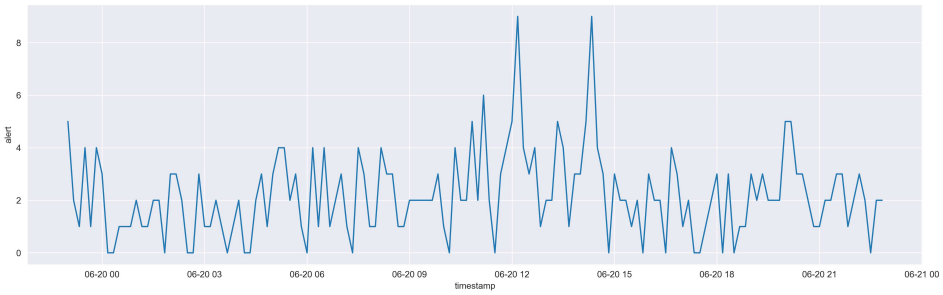


Fig. 6: All identified alerts accumulated

Figure 7 shows, in a heat map, a prominent line in the middle of the plot. This line indicates the countries involved in these important anomalies.

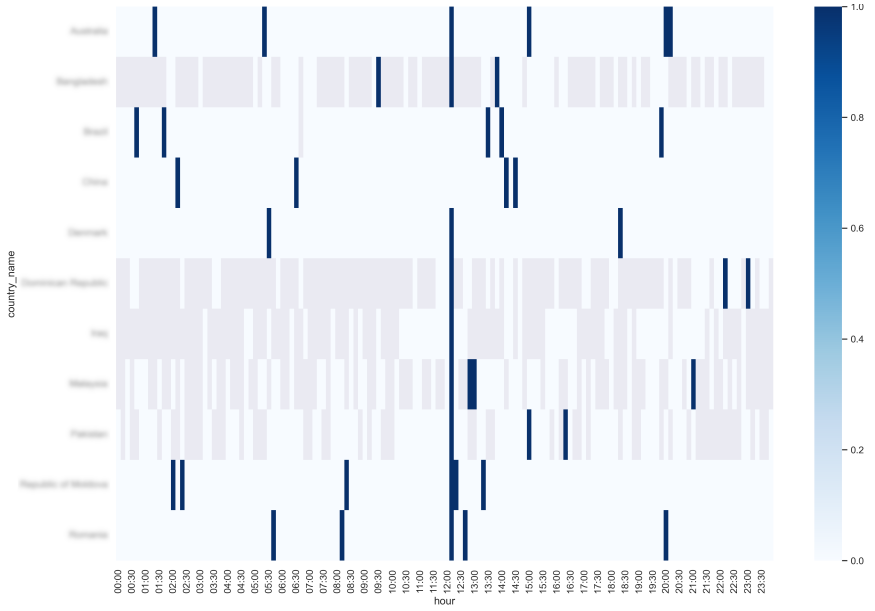


Fig. 7: Heat map of countries (sources) from which the requests originate

5 Prototype Multivariate E-S Method and its Architecture

In our previous prototype, we proposed our own method for time series analysis (see [Lu21]). We developed this original method to model the differently chosen minute-lags as features because established methods such as the Holt-Winters method were designed for univariate time series. In our optimized prototype, we took a step towards standardization by going back

to the well-established Holt-Winters method and making it able of handling multivariate time series. Listing 2 shows the corresponding code. This method is computationally intensive and therefore we anticipate running it in the cloud.

```

import datetime
begin_time = datetime.datetime.now()
mya = np.array([[0.15, 0.15], [0, 0.15]])
myb = np.array([[0.15, 0.15], [0, 0.15]])
myc = np.array([[0.15, 0.15], [0, 0.15]])

def multivariate_ES (data: np.array, h_period: int, A: np.array, B: np.array, C:
↪  np.array, m: int, year_size: int):
    L = []
    T = []
    data = data[-h_period:, :]
    inits = initialize (data,year_size)
    L.append (inits[0])
    T.append (inits[1])
    S = []
    for i in range (inits[2].shape[0]):
        S = np.append ([S], inits[2][i,:])
    shape = (int(S.shape[0]/2), 2)
    S = S.reshape(shape)
    s = []
    for i in range (inits[2].shape[0]):
        s = np.append ([s], inits[2][i,:])
    shape = (int(s.shape[0]/2), 2)
    s = s.reshape(shape)

    for i in range (h_period-1):
        L.append ((A @ (data[i+1,:] - S[i % year_size])) + ((np.identity(2) - A)
        ↪  @ (L[i]+T[i])))
        T.append ((B @ (L[i+1] - L[i])) + ((np.identity(2) - B) @ T[i]))
        S [(i+1)
        Smean = S.mean(axis = 0)
        for j in range(year_size):
            S [j] = S[j] - Smean
        s = np.append ([s], (S [(i+1) % year_size]))
        shape = (int(s.shape[0]/2), 2)
        s = s.reshape(shape)
    mves = []
    for i in range(6):
        mves.append(L[-1] + (i+1)*T[-1] + S[(h_period-1) % year_size])
    return pd.DataFrame(L), pd.DataFrame(T), pd.DataFrame(s), mves
```

Listing 2: Our Multivariate E-S Method

6 Conclusions and Future Research

In this paper, we describe a time series method and a cloud-native architecture for detecting DDoS attacks - which are one particularly important type of IT infrastructure anomalies in companies today. Our prototype was trialed in the data centers of Germany's Federal Employment Agency (FEA), and we used insights from its operation over the past year to optimize it and plan future refinements.

One area of improvement is the alert capability of the system. We plan to implement a live/real-time alert. This will alert users not only based on retrospective analysis, but also by detecting anomalies during operation and flagging them immediately, as well as by forecasting future anomalies. Automatic report generation will also be included along other commodity functions. In addition, we plan to differentiate more precisely which alert we are dealing with (compared to the current alert function which just identifies a generic alert). The system should be able to determine if an anomaly is really DDoS-related or rather something due to internal service level losses. We can achieve this by further refining the feature engineering. For example, response time is more of a service level feature and worth considering if that should be included for DDoS detection.

Another area of improvement is developing a true real-time architecture. We have demonstrated the Azure prototype in Section 3. The next step is to conceptualize and implement the full data pipeline into Azure such that computation and identification of anomalies will be completed in (almost) real-time. A follow-up step is to optimize the sample rate and, if necessary, define the reference time series for evaluation of individual measurements. Yet another future step is the conversion to real-time using containers according to FEA standards. We also plan to check raw data for usability (classifying whether evaluation is possible). Moreover we want to automatically perform analysis on any elevated anomaly density and document it, in order to support rapid root cause analysis. Additional improvements include classification of anomalies (e.g. technical failure external, internal, crawler access, etc.) and supporting the classification by deriving further metrics from the log data and analyzing them.

This work contributes to the growing body of literature regarding anomaly detection for IT infrastructure, in general, and DDoS attacks, in particular. We anticipate a continuing need for research in these areas to refine the existing anomaly detection methods and systems and support and protect increasingly complex IT infrastructures for digital transformation, internet of things, and other emerging IT trends.

Literatur

[Fe19] Fernandes, G.; Rodrigues, J. J.; Carvalho, L. F.; Al-Muhtadi, J. F.; Proença, M. L.: A comprehensive survey on network anomaly detection. Telecommunication Systems 70/3, S. 447–489, 2019.

[GG20] Gjorgiev, L.; Gievska, S.: Time Series Anomaly Detection with Variational Autoencoder Using Mahalanobis Distance. In (Dimitrova, V.; Dimitrovski, I., Hrsg.): ICT Innovations 2020: ICT Innovations 2020. Machine Learning and Applications. Bd. 1316. Communications in Computer and Information Science, Springer, S. 42–55, 2020, ISBN: 978-3-030-62097-4.

[He20] He, Q.; Zheng, Y. J.; Zhang, C.; Wang, H. Y.: MTAD-TF: Multivariate Time Series Anomaly DetectionUsing the Combination of Temporal Pattern and Feature Pattern. Complexity 2020/, 2020.

[Jo66] Jones, R. H.: Exponential Smoothing for Multivariate Time Series. Journal of the Royal Statistical Society 28/1, S. 241–251, 1966.

[KD18] Kotu, V.; Deshpande, B.: Data Science: Concepts and Practice. Morgan Kaufmann, 2018, ISBN: 978-0-12-814761-0.

[Li20] Li, J.; Izakian, H.; Pedrycz, W.; Jamal, I.: Clustering-based anomaly detection in multivariate time series data. Applied Soft Computing Journal 100/, 2020.

[LPJ17] Li, J.; Pedrycz, W.; Jamal, I.: Multivariate Time series Anomaly Detection: A Framework of Hidden Markov Models. Applied Soft Computing 60/, S. 229–240, 2017.

[Lu21] Ludsteck, J.; Sultanow, E.; Chircu, A.; Herget, G.; Seßler, M.: An architecture for detecting infrastructure anomalies at Germany's Federal Employment Agency. In: INFORMATIK 2021 – Computer Science & Sustainability. Gesellschaft für Informatik, 2021.

[NE22] NETSCOUT: Threat Intelligence Report Issue 8: Findings from 2nd Half 2021./, 2022.

[PA89] Pfeffermann, D.; Allon, J.: Multivariate exponential smoothing: Method and practice. International Journal of Forecasting 5/1, S. 83–98, 1989.

[Zh20] Zhao, H.; Wang, Y.; Duan, J.; Huang, C.; Cao, D.; Tong, Y.; Xu, B.; Bai, J.; Tong, J.; Zhang, Q.: Multivariate Time-Series Anomaly Detection via Graph Attention Network. In: 2020 IEEE International Conference on Data Mining (ICDM). IEEE, 2020.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1195

Flexibilität in Cloud ERP Systemen für bessere Unternehmensprozesse

Ansatzpunkte für individuelle Anpassungen

Dirk Bamberger[1], Farid Okhovat[2], Dr. Carsten Brockmann[3], Martina Sturm[4]

Abstract: adesso hat mit der New School of IT das Paradigma, die Vorteile des Cloud-Einsatzes voll auszuschöpfen ohne die Nachteile in Kauf nehmen zu müssen. Im Rahmen des Auswahlprozesses eines neuen ERP-Systems stand früh fest, dass Cloud Lösungen favorisiert werden. Klassische Vorbehalte gegenüber Cloud Lösungen wie Sicherheit, etc. gab es bei adesso nicht. Alle Anbieter mussten aufzeigen, wie die Funktionalität und Erweiterbarkeit des entsprechenden Systems ausgeprägt ist.
Nach der Entscheidung für S/4HANA Public Cloud (S4HC) begann die Übertragung und Anpassung der eigenen Prozesse auf die Standardabläufe im System wie auch der Umbau der Betriebsorganisation. Bei der Umgestaltung der eigenen Prozesse wurde schnell klar, dass es in S4HC nicht nur einen Standard-Weg, sondern auch verschiedene Optionen gibt. Ebenso besteht die Möglichkeit für Erweiterungen durch Partnerprodukte oder Eigenentwicklungen im SAP Technologiestack.

Keywords: Cloud ERP-Systeme, Prozessflexibilität, Activate

1 Einleitung

adesso wurde 1997 gegründet und hat seitdem ein sehr starkes Wachstum verzeichnet. So sind gegenwärtig 6.300 Mitarbeiter an 56 Standorten tätig, generieren einen Umsatz von fast 700 Millionen Euro. Der Erfolg der Muttergesellschaft wie auch der Tochtergesellschaften ist durch Flexibilität und starkes Wachstum geprägt, weshalb ein ERP-System gewählt wurde, dass diese Anpassungen leichtgewichtig zulässt. Mit der Einführung von S/4Hana Public Cloud sollen die Best Practices der SAP verwendet werden, ohne die Differenzierungsmarkmale von adesso zu verlieren. Hierbei ist ein starker Prozessfokus im Projekt notwendig, gepaart mit der technologischen Kompetenz der IT. Dies entspricht dem adesso Ansatz der New School of IT [Ad22].

[1] tabto consulting, Eschenweg 17, 35708 Haiger, dirk.bamberger@tabto.de
[2] adesso orange AG, Managing Consultant, Herriotstraße 1, 60528 Frankfurt am Main, Okhovat, Farid farid.okhovat@adesso-orange.com
[3] SAP Deutschland SE & Co. KG, Senior Project Manager S/4HANA Public Cloud, Rosenthaler Str. 30, 10178 Berlin, carsten.brockmann@sap.com
[4] SAP Deutschland SE & Co.KG, Delivery Manager S/4HANA Public Cloud, Frankfurter Str.1, 65760 Eschborn, Martina.Sturm@sap.com

doi:10.18420/inf2022_102

Um die Ziele zu erreichen, wurde die SAP Methodik Activate angewandt und projektspezifisch erweitert. So wurden beispielsweise das Backlog in einem Ticketsystem und wichtige inhaltliche Informationen in einem Projektmanagementtool aus dem agilen Umfeld erfasst.

2 Darstellung der Implementierungsmethode ACTIVATE

SAP Activate ist eine Kombination aus SAP Best Practices, einer Methodik und geführter Konfiguration:

- **SAP Best Practices**: SAP Best Practices für SAP S/4HANA Public Cloud bieten einsatzfertige digitalisierte Geschäftsprozesse für Finanzwesen und Logistik, die in andere Cloud-Lösungen wie SuccessFactors Employee Central oder Concur integriert sind.
- **Geführte Konfiguration:** Die Konfiguration mit Benutzerführung vereinfacht die Einführung von SAP S/4HANA. Sie steht durchgehend von der anfänglichen Einrichtung bis zu kontinuierlichen Innovationen nach dem Produktivstart zur Verfügung. Fachbereiche und die IT-Abteilung sind dadurch in der Lage zusammenzuarbeiten.
- **Durchgängige Methodik:** Die Methodik von SAP Activate ist der Ausgangspunkt für jede Einführung von SAP S/4HANA. Sie ist der Nachfolger der Methoden ASAP und SAP Launch. Kern der Methodik ist die Scope Item basierte Einführung des ERP-Systems. Ein im System vorhandener Prozess wird als Scope Item bezeichnet. Der Funktionsumfang der Scope Items variiert von kleinen, klar abgegrenzten Schritten bis hin zur Abrechnung von Projekten. Die Aktivitäten je nach Phase können der Abbildung 1 entnommen werden.

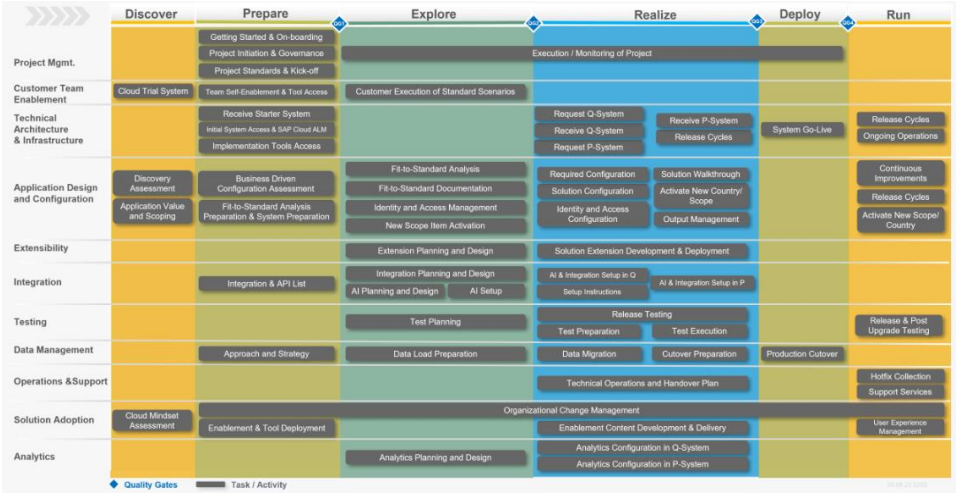


Abbildung 1: ACTIVATE Artefakte bei der S/4HANA Public Cloud

Mittels der drei Komponenten erfolgt die gezielte und strukturierte Einführung des Cloud ERP-Systems.

3 Anpassungsmöglichkeiten mit Cloud-Systemen

Im Rahmen der Explore Phase gab es einen sehr starken Schwerpunkt auf die Unternehmensprozesse. Jedem Prozessschritt wurde das relevante Scope Item zugeordnet, um die Implementierung zu ermöglichen. Das Vorgehen ist eingebettet in eine Projekt-Governance mit unterschiedlichen Rollen.

3.1 Prozessgestaltung und Optionen

Im Rahmen der ERP-Einführung werden die einzelnen Prozesse und relevanten Systeme im Soll erfasst. Durch die Ende-zu-Ende Betrachtung wird beispielsweise sichergestellt, dass Aufträge, die im CRM-System angelegt werden, ihren Weg in die Projektverwaltung des S/4HANA Public Cloud finden.

Im Rahmen des Workshops wurden zunächst die Best Practice Prozesse vorgestellt. Es erfolgte die Überlegung der adesso-Mitarbeiter, welche Bestandteile der aktuellen Prozesse unabdingbar in S/4HANA Public Cloud vorhanden sein müssen. Mittels einer intensiven Moderation wurde geprüft, ob die Anforderungen einen tatsächlichen Mehrwert darstellen. Sobald der Mehrwert erkennbar war, galt es die Frage zu beantworten, wie er zu realisieren ist. Hierfür gab es zwei grundsätzliche Ansatzpunkte:

- **Unterschiedliche Wege durch das System führen zum gewünschten Ergebnis**. Beispielhaft sollte ein bestimmtes Dashboard Einstieg in die Transaktionen für die Sachbearbeiter sein. Diese Darstellung ließ sich nicht über den direkten Prozessweg in der gewünschten Art lösen. Stattdessen wurde ein embedded SAP Analytics Cloud (SAC) Dashboard aufgebaut, von dem aus die Sachbearbeiter in die einzelnen Fälle abspringen können.
- **Ergänzung von Feldern.** Durch die Ergänzung von Feldern in Tabellen konnten adesso-spezifische Merkmale hinzugefügt und in der weiteren Logik mit berücksichtigt werden.
- **Spezielle Darstellungen:** Mittels eigen zusammengestellter Tabellen können Daten in einer andersartigen Form dargestellt werden und die Bearbeitung zu vereinfachen. Ein Beispiel ist die Zusammenstellung von Projektrelevanten Daten, die an die adesso Eigenentwicklung fürs operative Projektmanagement übergeben werden.
- **Ergänzungen von Workflows:** Durch die Ergänzung von Workflows können neue Bearbeitungsschritte hinzugefügt werden. Beispielhaft soll bei der Anlage eines neuen Lieferanten die Kontonummer im vier-Augen Prinzip geprüft werden, was zu einer Ergänzung des bestehenden Workflows führt.
- **In-App Development (perspektivisch):** Zukünftig wird es möglich sein, Funktionalität in S4HC selbst und nicht über das Ökosystem zu entwickeln. Vorteil ist ein leichterer Umgang mit dem Wartungszyklus. Ebenso werden Länderlokalisierungen hierüber abgebildet werden können.

3.2 Nahtlose Erweiterungsmöglichkeiten innerhalb des SAP Ökosystems

Manche Anforderungen waren derart komplex, dass weder unterschiedliche Wege noch weitere Datenfelder den Sachverhalt auflösen konnten. Für solche Fälle existiert eine große Bandbreite an Veränderungsmöglichkeiten innerhalb des SAP-Ökosystems [Br14].

Dreh- und Angelpunkt hierfür ist die Business Technology Plattform, die es neben der Erstellung von einfachen Workflows ermöglicht, Anwendungen vollständig selber zu entwickeln.

adesso hat hiervon Gebrauch gemacht, und eine Lösung für das operative Projektmanagement z.B. für die Stundenerfassung selber entwickelt.

Perspektivisch wird die Drei-Systemlandschaft (3SL) im Cloudbereich die Möglichkeit bieten, auf einem Entwicklungssystem Dinge auszuprobieren, die dann auf der Qualitätsumgebung weiter fortentwickelt und getestet werden um anschließend in der Produktivumgebung verwendet zu werden.

3.3 Anbindung von Drittsystemen

Sollte die benötigte Funktionalität weder aus dem ERP selbst noch eigen entwickelt werden, besteht die Möglichkeit zur Anbindung von Drittsystemen an das SAP-Umfeld. Um bereits die Drittanwendungen schnell anzubinden, werden verschiedene Konnektoren bereit gestellt. So wurde beispielsweise das CRM-System eines anderen Herstellers mittels des Konnektors angebunden.

4 Program Governance mit klarem inhaltlichem Fokus

Eine ERP-System Einführung wie bei adesso mit der erklärten Zielsetzung der Optimierung der Unternehmensprozesse ist immer auch und zuallererst eine umfangreiche Transformation und damit Change Management Aufgabe. Um diese erfolgreich zu meistern, sind viele Menschen im Unternehmen einzubinden und es gilt, deren Zusammenwirken in Form einer funktionierenden Governance-Struktur zu verabreden.

Um die Prozessgestaltung und deren Bewertung hinsichtlich Veränderungsumfang und -realisierbarkeit mit den relevanten Beteiligten und effizient diskutieren und entscheiden zu können, braucht es unter Anderem klar definierte und wahrgenommene Verantwortlichkeiten für eine funktionieren Governance im Programm. Wir haben daher von Anfang an eine Trennung der Verantwortung für die fachlichen Fragestellungen und Entscheidungen von der methodischen und steuernden Verantwortung etabliert - nicht zuletzt, um den Teilnehmern aus den Fachbereichen eine klare Fokussierung auf die Inhalte zu ermöglichen.

Die Projektsteuerung und methodische Leitung übernehmen in den Teams die Scrum Master, die durch ihre Rolle die Fachbereiche als „Servant Leader“ unterstützen und entlasten.

Alle zu realisierenden Prozessanforderungen und die daraus resultierenden Aufgaben bei der Realisierung der Lösung werden entsprechend über Backlogs gemanagt, die in der Regel und gemäß agiler Standardmethoden durch sog. Product Owner verantwortet werden. Eine einzige Rolle reichte zur Steuerung für das umfangreiche Transformationsprogramm bei adesso nicht aus. Vielmehr erforderten der Umfang der fachlichen Themenstellungen, der unterschiedliche Grad an benötigter Prozess- und Lösungskompetenz je Softwareprodukt und Art der Themen im Backlog und letztlich auch die Tragweite der involvierten Entscheidungen die Ausprägung dieser Rolle mit unterschiedlichen Schwerpunkten und aus unterschiedlichen Hierarchiestufen aus der Organisation.

Für die fachliche Verantwortung der betroffenen Geschäftsbereiche wurden daher verschiedene Rollen festgelegt. Die Rollen lassen sich wie folgt charakterisieren:

- *Business Owner* sind die für einen Fachbereich final verantwortlichen Entscheider aus der Linienorganisation, also z.B. die Leiter:innen HR oder Finance
- *Portfolio Owner* verantworten jeweils ein inhaltlich abgeschlossenes, großes Themenfeld (bspw. Finance, HR, Project Management)
- *Product Manager* verantworten Unterbereiche (bspw. ist der Einkauf im Bereich Finance verankert)
- *Product Owner* arbeiten für die Unterbereiche fachspezifische Anforderungen aus und verantworten damit konkrete Artefakte im Rahmen der zu realisierenden Backlogs (bspw. einen Freigabeworkflow im Einkauf, ein Datenmigrationsobjekt, ein Rechnungsformular, etc.).



Abbildung 2: Darstellung der Programmrollen der Fachbereiche

Darüber hinaus tragen zu einer erfolgreichen Zusammenarbeit im Programm weitere Rollen bei, die meist durch das IT-Team ausgefüllt werden und dabei bspw. Querschnittsthemen wie Schnittstellenrealisierungen und das Ausgestalten der Formulare oder auch die Datenmigration federführend vorantreiben und umsetzen.

Auch eine gut durchdachte Integration der Consulting- und Realisierungs-Rollen ist für eine funktionierende Program Governance unverzichtbar, die in vielen Fällen und wie auch in unserem Programm durch Implementierungspartner wahrgenommen werden.

Gerade im Rahmen solcher Einführungen von standardisierten Public Cloud Lösungen existieren noch einmal gestiegene Anforderungen an Transformationswillen und -fähigkeit der Organisation auf der einen Seite und andererseits auch der Bedarf an Kompetenzen, die vorhandenen Möglichkeiten der flexiblen Lösungsgestaltung intelligent auszuschöpfen. Insgesamt verdeutlicht dies auch die zunehmende Bedeutung einer gut durchdachten, auf die individuellen Rahmenbedingungen des Unternehmens zugeschnittenen und funktionierenden Program Governance. Ein effizientes

Zusammenwirken vieler verteilter Kompetenzen und Fähigkeiten kann und sollte auf diese Weise sichergestellt werden.

5 Fazit

Die Einführung von Cloud Software ist dann erfolgreich, wenn der Fachbereich des Kunden bereit ist, sich von bestehenden Prozessen zu lösen, offen für die standardisierte und pragmatische Vorgehensweisen ist und proaktiv in dem Transformationsprozess mitarbeitet. Wir haben daher alle Stakeholder frühzeitig und umfassend in das Implementierungsprojekt mit eingebunden.

Eine weitere wesentliche Rolle spielt der Implementierungspartner SAP. Dieser hat uns beispielsweise im Rahmen der Fit-to-Standardworkshops den Raum geben, die Prozessanforderungen darzulegen und zu diskutieren. Die Einschätzungen der Stakeholder waren teilweise unterschiedlich.

Die Konflikte dabei müssen durch einen gut moderierten Dialog aufgelöst werden, der sowohl Prozessziele und -ergebnisse als auch technische Spielräume einschließt. Letztlich ist auch eine klare Überzeugung aller Entscheider für die Vorteile eines standardisierten Ansatzes unabdingbar.

Unter Berücksichtigung von Standardisierung, Integration und Automatisierung wurden Umsetzungskonzepte erstellt und gemeinsam evaluiert. Die Konzepte wurden an standardisierte Tools, cloudbasierte Plattformen und Services ausgerichtet, so dass notwendige Individualisierung umgesetzt werden konnten.

Der Dienstleister muss daher handwerklich gut aufgestellt sein, um die Optionen der Individualisierung, die in einem S/4HANA Public Cloud Architektur gegeben sind, zu designen und in die Gesamtsystemarchitektur zu integrieren.

6 Literaturverzeichnis

[Ad22] adesso, New School of IT, IT muss neu gedacht werden, https://www.adesso.de/de/impulse/new-school-of-it, Stand: 11.07.2022.

[Br14] Brockmann, An approach to design the business model of an ERP vendor. GITO, Berlin (2014)

[Sa22] SAP, SAP Activate, https://www.sap.com/germany/products/activate-methodology.html, Stand: 11.07.2022.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1203

Entwicklung einer Open-Data-Referenzarchitektur für die Luftfahrtindustrie

Ann-Kathrin Losse[1], Melanie Gehrke[2], André Ullrich[3], Christian Czarnecki[4], Eldar Sultanow[5], Carsten Breithaupt[6], Christian Koch[7]

Abstract: Open Data impliziert die freie Zugänglichkeit, Verfügbarkeit und Wiederverwendbarkeit von Datensätzen. Obwohl hochwertige Datensätze öffentlich verfügbar sind, ist der Zugang zu diesen und die Transparenz über die Formate nicht immer gegeben. Dies mindert die optimale Nutzung des Potenzials zur Wertschöpfung, trotz der vorherrschenden Einigkeit über ihre Chancen. Denn Open Data ermöglicht das Vorantreiben von Compliance-Themen wie Transparenz und Rechenschaftspflicht bis hin zur Förderung von Innovationen. Die Nutzung von Open Data erfordert Mut und eine gemeinsame Anstrengung verschiedener Akteure und Branchen. Im Rahmen des vorliegenden Beitrags werden auf Grundlage des Design Science-Ansatzes eine Open Data Capability Map sowie darauf aufbauend eine Datenarchitektur für Open Data in der Luftfahrtindustrie an einem Beispiel entwickelt.

Keywords: Open Data; Unternehmensarchitekturen; Referenzarchitekturen; Luftfahrt

1 Einleitung

Die Umsetzung von Verbesserungspotentialen und innovativen Geschäftsmodellen wird vom Vorhandensein qualitativ hochwertiger Daten beeinflusst [HV11; Zh19]. In diesem Zusammenhang wird der offene Zugriff auf Daten - auch als Open Data bezeichnet - als ein wichtiger Treiber angesehen [LN21; Mc13; SSS+20]. Die Idee von Open Data ist die Bereitstellung und der Zugriff auf jederzeit frei verfügbare, wiederverwendbare Datensätze [MS16]. Dennoch können die Potenziale von Open Data noch nicht immer ausgeschöpft werden. Zwar existiert eine Vielzahl an Plattformen, welche die Daten anbieten, doch ist die Integration in die Geschäftsabläufe und die bestehende IT-Landschaft häufig unklar auf Seite der Unternehmen. Das liegt unter anderem an sich verändernden notwendigen Kompetenzen der Architekten [Ul21] und betrifft auch die Verfahren, um Daten für den eigenen Gebrauch zu transformieren, sowie die Echtzeitverarbeitung der Datenflüsse. In diesem Zusammenhang werden Referenzarchitekturen als mögliche Lösung angesehen [CD17; Lo18; MS16]. Als Anwendungskontext von Open Data werden u.a E-Government

[1] Capgemini, Potsdamer Platz 5, 10785 Berlin, Deutschland, ann-kathrin.losse@capgemini.com
[2] Capgemini, Mainzer Landstraße 178, 60327 Frankfurt, Deutschland, melanie.gehrke@capgemini.com
[3] Universität Potsdam, Deutschland, andre.ullrich@wi.uni-potsdam.de
[4] FH Aachen, Eupener Str. 70, 52066 Aachen, Deutschland, czarnecki@fh-aachen.de
[5] Capgemini, Bahnhofstraße 30, 90402 Nürnberg, Deutschland, eldar.sultanow@capgemini.com
[6] Lufthansa, Deutschland, carsten.breithaupt@dlh.de
[7] Technische Hochschule Nürnberg, Kesslerpl. 12, 90489 Nürnberg, Deutschland, christian.koch@th-nuernberg.de

doi:10.18420/inf2022_103

[MO11], Smart Cities [Sc17] sowie der Dienstleistungssektor [JCZ12] vorgeschlagen. Auch in der Luftfahrtindustrie sind erste Fallbeispiele dokumentiert [SSS19; US22]. Dieser Beitrag präsentiert vorläufige Ergebnisse einer Referenzarchitektur für Open Data für die Luftfahrtindustrie und ist Bestandteil eines umfassenden Forschungsvorhabens, dessen Ziel die Entwicklung einer Open-Data-Referenzarchitektur für den Dienstleistungssektor ist. Beteiligt daran ist eine interdisziplinäre Gruppe aus Wissenschaft, Industrie und Beratung.

In Abschnitt 2 wird das Vorgehen bei der Entwicklung der Referenzarchitektur beschrieben. Erste Entwicklungsergebnisse werden in Abschnitt 3 vorgestellt, wobei der Fokus auf einer Capability Map liegt. Außerdem wird in Abschnitt 4 ein erster Anwendungsfall von Open Data an einem konkreten Industriebeispiel veranschaulicht. Der Beitrag schließt mit einem Fazit und Ausblick.

2 Methodik und Vorgehen

Als Leitfaden zur Entwicklung der Referenzarchitektur dient der Design Science Research (DSR) Ansatz [He04; Pe07]. Kennzeichnend ist die iterative Erarbeitung eines Artefakts und dessen Bewertung. Die Praxiseignung und Anwendbarkeit des Artefakts sollen in einer für das Artefakt relevanten Umgebung demonstriert werden. Die Forschung strukturiert sich anhand des DSR-Publikationsschemas [He04]: (1) Problembeschreibung, (2) Entwurfsmethode, (3) Artefaktbeschreibung, (4) Artefaktbewertung sowie (5) Diskussion und Schlussfolgerung.

Das durch diesen Ansatz entstehende Artefakt ist eine Referenzarchitektur für Open Data. Üblicherweise kommen Referenzarchitekturen in der (Wirtschafts-)Informatikdisziplin zum Einsatz [CD17; WF06]. Eine Referenzarchitektur unterscheidet sich von einer konkreten Architektur insbesondere hinsichtlich ihrer Wiederverwendbarkeit und einer Allgemeingültigkeit für eine Bandbreite von Anwendungsfällen [FL07]. Die in diesem Beitrag vorgestellte Referenzarchitektur wurde durch reale Projekte in Kooperation mit einem Unternehmen aus der Luftfahrtindustrie verprobt. Sowohl die Modellierung der Referenzarchitektur, als auch die konkreten Anwendungen dieser auf die beiden Branchen Luftfahrt und Bankwesen können als Design Artefakte angesehen werden. Zum Zeitpunkt der Verfassung dieses Beitrags sind die beiden Anwendungen noch in der Bearbeitung und Evaluierung in Kooperation mit den Praxispartnern.

In Bezug auf das von [GH13] eingebrachte DSR Knowledge Contribution Framework ist die Entwicklung der Open-Data-Referenzarchitektur im Verbesserungsquadranten anzusiedeln. Dieser zielt auf die Entwicklung neuer Lösungen für bekannte Problemdomänen ab. Zwar sind Referenzarchitekturen für sich ein gut erforschtes Thema, allerdings sind Lösungen für die Anwendung von Open Data bisher rar, jedoch zeitgleich verstärkt nachgefragt und benötigt. Um auf diesen Bedarf einzugehen, setzt diese Forschung auf die enge Einbindung der Praxis. Im Anwendungsfall waren die Co-Autoren des Beitrags in führenden Rollen an den Projekten beteiligt. Somit basiert der zugrundeliegende Forschungsansatz zur

Ergründung der Praxisbedarfe im Bereich Open Data auf direkten Beobachtungen und deren Dokumentation.

3 Vorläufige Ergebnisse bei der Entwicklung der Referenzarchitektur

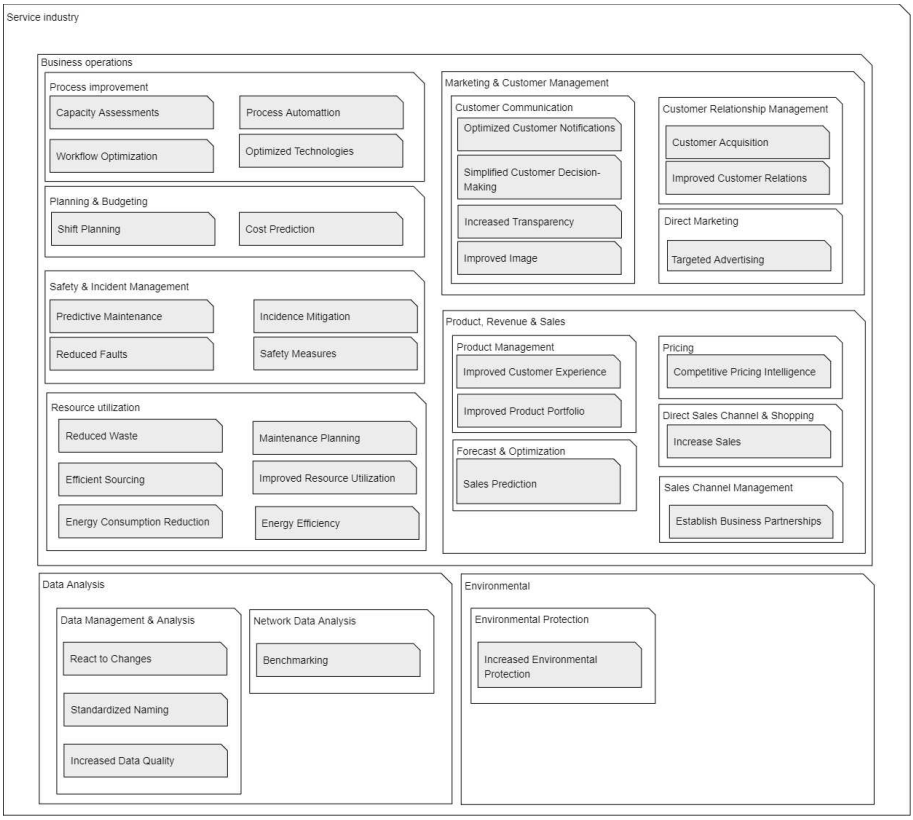


Abb. 1: Vorläufige Open Data Capability Map

Die Struktur der entwickelten Referenzarchitektur orientiert sich an den Standards des Open Group Architecture Framework (TOGAF):

- Die **Geschäftsarchitektur** (Business Architecture) enthält den Geschäftsprozess sowie eine Zuordnung der einzelnen Geschäftsfunktionen (Business Capabilities). Diese werden in Bezug auf Open Data klassifiziert und möglichen Datenquellen zugeordnet.
- Die **Datenarchitektur** umfasst die Datenobjekte und deren Beziehungen. Dabei werden die für Open Data relevanten Datenobjekte identifiziert.

- **Applikationsarchitektur** stellt den Zusammenhang zwischen den geschäftlichen Funktionen, den Datenobjekten und der technischen Umsetzung dar. Sie definiert die notwendigen technischen Komponenten.

Aus inhaltlicher Sicht werden die für Open Data relevanten Architerkturelemente in einer Open Data Capability Map zusammengefasst. Diese stellt eine hierarchische Strukturierung und Klassifizierung der notwendigen Funktionen dar. Damit liefert sie eine erste Referenz, die als Ausgangspunkt für die offene Erfassung, Bearbeitung und Austausch von Daten genutzt werden kann. Das vorläufige Ergebnis der Open Data Capability Map ist in Abb. 1 zusammengefasst.

4 Beispielhafte Anwendung der Referenzarchitektur bei der Lufthansa

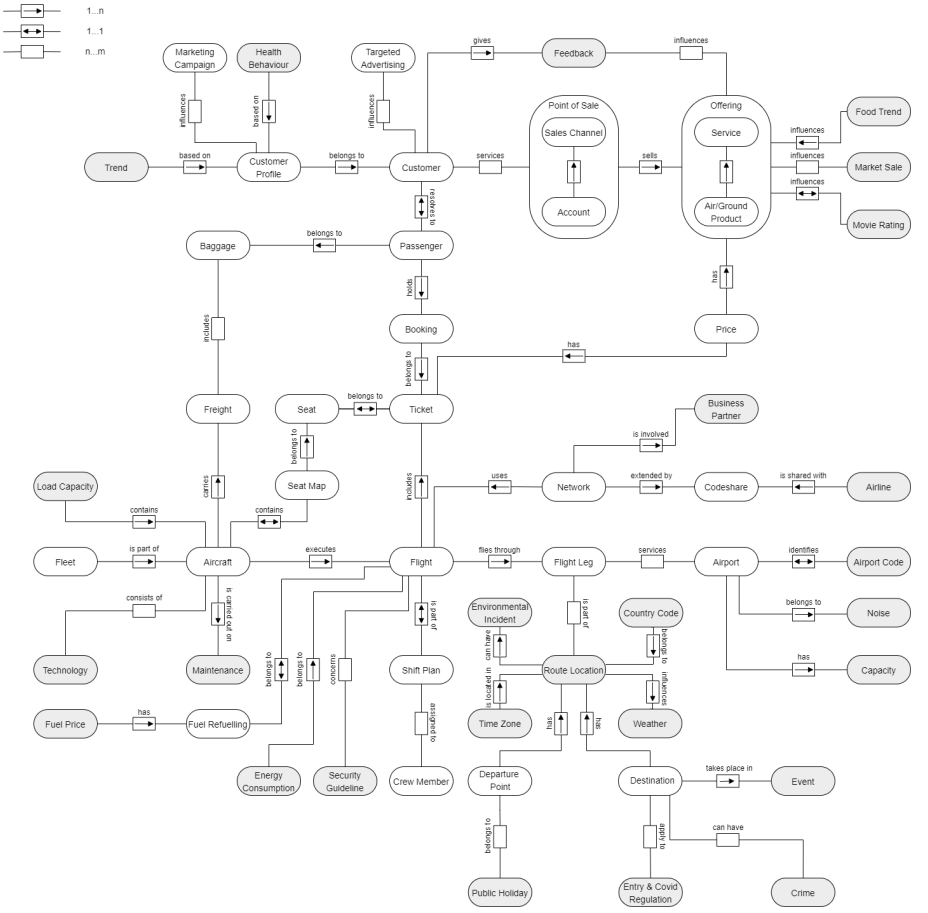


Abb. 2: Konkrete Ausprägung der Datenarchitektur bei Lufthansa

Die im vorherigen Abschnitt vorgestellte Open Data Capability Map wurde in Kooperation mit der Lufthansa in einem konkreten Praxisprojekt angewendet.

Im Rahmen von Interviews mit Experten der Lufthansa und durch die Bereitstellung von Unterlagen, konnte ausgehend von der allgemeinen Open Data Capability Map (Abb. 1) eine auf die Lufthansa zugeschnittene Capability Map erarbeitet werden. Weitergehend wurden in einem virtuellen Workshop konkrete, öffentliche Datenquellen identifiziert, die zur Erweiterung und Verbesserung der Geschäftsprozesse der Luftfahrt dienen können. Basierend auf den Workshop-Resultaten, war die Forschergruppe in der Lage, die in Abb. 2 veranschaulichte Datenarchitektur zu entwickeln und so die konkreten Anknüpfungspunkte für Open Data Quellen zu visualisieren. Eine Validierung der bisher erarbeiteten Artefakte ist für einen weiteren virtuellen Workshop geplant.

5 Fazit und Ausblick

Der offene Austausch von Daten kann zukünftig ein wichtiger Treiber für Innovationen und Geschäftsmodelle sein. Dazu ist ein gemeinsames Verständnis auf Geschäfts-, Daten- und Applikationsebene unerlässlich. Als Lösung wird dazu die Entwicklung einer Open Data-Referenzarchitektur vorgeschlagen. Diese Entwicklung wird aktuell von einer Forschungsgruppe bestehend aus Wissenschaft, Industrie und Beratung vorangetrieben. In einem ersten Schritt wurde eine Referenzarchitektur für die Luftfahrtindustrie entwickelt und in einem Praxisprojekt bei der Lufthansa angewendet.

Der Beitrag zeigt vorläufige Ergebnisse dieses Forschungsvorhabens. Die Open Data Capability Map liefert eine mögliche inhaltliche Strukturierung als Startpunkt für die Umsetzung von Open Data. Aktuell werden die bestehenden Artefakte weiter detailliert und mit Industrievertretern diskutiert. Darüber hinaus ist die Abstraktion und Übertragung auf weitere Industrien in Arbeit. Denkbar wäre eine Anwendung in der Finanzbranche, um die Relevanz der Referenzarchitektur in einer weiteren Dienstleistungsbranche zu verproben. Das vorliegende Praxisbeispiel der Lufthansa zeigt eine erste Validität. Für eine umfassende Evaluation sind jedoch weitere Praxisprojekte notwendig.

Literatur

[CD17] Czarnecki, C.; Dietze, C.: Reference Architecture for the Telecommunications Industry: Transformation of Strategy, Organization, Processes, Data, and Applications. Springer International Publishing, Cham, 2017, ISBN: 978-3-319-46757-3.

[FL07] Fettke, P.; Loos, P.: Perspectives on reference modeling. In: Reference modeling for business systems analysis. IGI Global, S. 1–21, 2007.

[GH13] Gregor, S.; Hevner, A.: Positioning and Presenting Design Science Research for Maximum Impact. MIS Quarterly 37/, S. 337–356, Juni 2013.

[He04] Hevner, A.; R, A.; March, S.; T, S.; Park; Park, J.; Sudha, R.: Design Science in Information Systems Research. Management Information Systems Quarterly 28/, März 2004.

[HV11] Huijboom, N.; Van den Broek, T.: Open data: an international comparison of strategies. European journal of ePractice 12/1, S. 4–16, 2011.

[JCZ12] Janssen, M.; Charalabidis, Y.; Zuiderwijk, A.: Benefits, adoption barriers and myths of open data and open government. Information systems management 29/4, S. 258–268, 2012.

[LN21] Lnenicka, M.; Nikiforova, A.: Transparency-by-design: What is the role of open data portals? Telematics and Informatics 61/, S. 101605, 2021.

[Lo18] Loenen, B. v.; Vancauwenberghe, G.; Crompvoets, J.; Dalla Corte, L.: Open data exposed. In: Open data exposed. Springer, S. 1–10, 2018.

[Mc13] McKinsey & Company: Open data: Unlocking innovation and performance with liquid information, 2013, URL: https://www.mckinsey.com/~/media/mckinsey/business%20functions/mckinsey%20digital/our%20insights/open%20data%20unlocking%20innovation%20and%20performance%20with%20liquid%20information/mgi_open_data_fullreport_oct2013.pdf, Stand: 27. 11. 2021.

[MO11] Machado, A. L.; de Oliveira, J. M. P.: DIGO: An open data architecture for e-government. In: 2011 IEEE 15th international Enterprise distributed object computing conference workshops. IEEE, S. 448–456, 2011.

[MS16] Monino, J.-L.; Sedkaoui, S.: Big data, open data and data development. John Wiley & Sons, 2016.

[Pe07] Peffers, K.; Tuunanen, T.; Rothenberger, M. A.; Chatterjee, S.: A design science research methodology for information systems research. Journal of management information systems 24/3, S. 45–77, 2007.

[Sc17] Schieferdecker, I.; Tcholtchev, N.; Lämmel, P.; Scholz, R.; Lapi, E.: Towards an open data based ICT reference architecture for smart cities. In: 2017 Conference for E-Democracy and Open Government (CeDEM). IEEE, S. 184–193, 2017.

[SSS+20] Sundaram, R.; Sharma, D.; Shakya, D. et al.: Digital transformation of business models: a systematic review of impact on revenue and supply chain. International Journal of Management 11/5, 2020.

[SSS19] Sultanow, E.; Stolz, A.; Sprinck, L.: Maschinelles Lernen in Java: Frameworks und Praxis. JavaSPEKTRUM/3, S. 22–26, 2019.

[Ul21] Ullrich, A.; Bertheau, C.; Wiedmann, M.; Sultanow, E.; Körppen, T.; Bente, S.: Roles, tasks and skills of the enterprise architect in the VUCA world. In: 2021 IEEE 25th International Enterprise Distributed Object Computing Workshop (EDOCW). IEEE, S. 261–270, 2021.

[US22] U.S. Department of Transportation: Consumer Airfare Report: Table 1 - Top 1,000 Contiguous State City-Pair Markets, https://data.transportation.gov/Aviation/Consumer-Airfare-Report-Table-1-Top-1-000-Contiguo/4f3n-jbg2, 2022, Stand: 12. 04. 2022.

[WF06] Winter, R.; Fischer, R.: Essential Layers, Artifacts, and Dependencies of Enterprise Architecture. http://www.alexandria.unisg.ch/Publikationen/671233/, Jan. 2006.

[Zh19] Zhu, Z.; Wulder, M. A.; Roy, D. P.; Woodcock, C. E.; Hansen, M. C.; Radeloff, V. C.; Healey, S. P.; Schaaf, C.; Hostert, P.; Strobl, P. et al.: Benefits of the free and open Landsat data policy. Remote Sensing of Environment 224/, S. 382–385, 2019.

Der EmiLe Computer Club (ECC): Informatik an Schulen spannend und interessant vermitteln

Programmieren lernen an einer Montessori-Schule: Der EmiLe Computer Club (ECC)

Steffi Rudel[1,2]

Abstract: Informatische Bildung an Schulen zu vermitteln ist heute wichtiger denn je. Gerade an Mittelschulen ist es jedoch oft nicht einfach, Schülerinnen und Schüler für diese Themen zu begeistern. Dazu kommt, dass an dieser Schulart häufig fachfremde LehrerInnen mit wenig Vorwissen Informatik unterrichten sollen. In diesem Beitrag wird der EmiLe Computer Club (ECC) vorgestellt, der den Fokus auf den Teilbereich des Programmierens legt. Im ECC wird an einer Montessori-Schule im Münchner Südosten seit dem Schuljahr 2020/21 Programmieren von der 3.-8. Klasse spielerisch und ungezwungen vermittelt. Es wird das Konzept des ECC vorgestellt und ausgewählte Beispiele aus der Praxis werden präsentiert. Darüber hinaus soll der Beitrag die Diskussion zur Verbindung der Montessori-Pädagogik mit der Ausbildung junger Menschen im Programmieren anregen.

Keywords: Programmieren; informatische Bildung; Montessori-Pädagogik

1 Programmieren und Montessori - passt das zusammen?

Seit langem wird von vielen Seiten gefordert, mehr Informatikunterricht in die Schulen zu bringen. Diese Forderung griff Ende 2021 auch die Kultusministerkonferenz (KMK) in der ergänzenden Empfehlung „Lehren und Lernen in der digitalen Welt“ auf [Ku21], [Ge22]. Ferner wird bezüglich der Chancengleichheit von Mädchen und Frauen in der Informatik immer wieder betont, dass der Unterricht in diesem Bereich möglichst früh starten sollte (z.B. [He15], [HG22]).

Je nach Bundesland ist Informatikunterricht inzwischen in einigen Schularten und Klassenstufen verpflichtend (eine Übersicht bietet [SHF21]). In den Mittelschulen im Bundesland Bayern ist beispielsweise ab der 5. Klasse eine Schulstunde (45 Minuten) pro Woche vorgeschrieben [St22]. Dieser Unterricht soll jedoch vorrangig nicht von FachlehrerInnen unterrichtet werden, sondern soll möglichst von den Klassenleitungen der Mittelschule abgedeckt werden [Mu22].

Um auch fachfremden Lehrkräften den Programmierunterricht (als Teil des Informatikunterrichts) zu erleichtern, wird in diesem Artikel der EmiLe Computer Club vorgestellt. Außerdem soll mit dem Beitrag die Diskussion zur Verbindung der

[1] Universität der Bundeswehr München, Professur für Wirtschaftsinformatik, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Steffi.Rudel@unibw.de
[2] EmiLe Montessorischule Neubiberg, Arastr. 2, 85579 Neubiberg, S.Rudel@emile-montessori.de

doi:10.18420/inf2022_104

Montessori-Pädagogik mit der Ausbildung junger Menschen im Programmieren angeregt werden.

2 Das Konzept des EmiLe Computer Clubs (ECC)

Die EmiLe Montessorischule ist eine staatlich genehmigte private Bildungseinrichtung im Münchner Südosten. Die Schule vereint Grund-, Mittel- und Fachoberschule unter einem Dach und ist als Ganztagsschule konzipiert. Nach eigenen Angaben lebt die EmiLe „[…] eine erprobte Verbindung aus den Prinzipien Maria Montessoris sowie moderner Arbeitsmethoden und Lerntechniken“ [Em22]. Ein besonderer Schwerpunkt liegt auf den Naturwissenschaften, die in der EmiLe nach dem Prinzip IM TUN (Informatik, Mathematik sowie Technik, Umwelt, Naturwissenschaften) umgesetzt werden.

Der EmiLe Computer Club (ECC) wurde von der Autorin im Schuljahr 2020/21 eingeführt. Ziel des ECC ist es, ähnlich wie von Armoni et al. [AG14] propagiert, Schülerinnen und Schüler (SuS) spielerisch und ungezwungen an das Programmieren heranzuführen. So sollen das Interesse für Informatik geweckt und kausale Denkstrukturen gefördert werden. Ein besonderer Fokus im ECC wird auf Mädchen gelegt.

Auf textuelle Programmierung wird im ECC bewusst verzichtet, stattdessen wird ausschließlich mit visueller, blockbasierter Programmierung gearbeitet. [3] Diese bietet den Vorteil, dass auch Programmier-Anfänger einen einfachen Einstieg finden und das Selbstbewusstsein der SuS durch schnelle Erfolgserlebnisse gestärkt werden kann [SC22].

Der ECC findet jeweils ein Halbjahr lang einmal wöchentlich am Nachmittag statt und ist aktuell über die Jahrgangsstufen drei bis acht aufeinander aufbauend konzipiert. Die SuS wählen sich jeweils zu Beginn eines Halbjahres für die Kurse ein, wobei Mädchen gezielt zur Teilnahme ermutigt werden.

Die folgende Tabelle zeigt die Kurse, Klassenstufen, Zeitansätze und Kompetenzerwartungen.

Kurs	**Klassenstufe**	**Zeitansatz / Woche**	**Kompetenzerwartungen**
ECC P1	3./4. Klasse	Zwei Schulstunden (90 Minuten)	Wir funktioniert Programmieren – und wie können wir uns selbst mit Software verwirklichen?

[3] Ein essentieller Bestandteil der Montessori-Pädagogik ist die Freiarbeit mit Materialien. Ein wichtiges Qualitätsmerkmal der Materialien ist dabei die „Isolierung einer einzigen Eigenschaft“ in jedem Material [Sc16]. Der bewusste Verzicht auf die textuelle und die Fokussierung auf die visuelle, blockbasierte Programmierung könnte als die Umsetzung dieses Qualitätsmerkmals interpretiert werden.

Kurs	Klassenstufe	Zeitansatz / Woche	Kompetenzerwartungen
ECC P2	5./6. Klasse	Zwei Schulstunden (90 Minuten)	Wie verändert Software unsere Welt – und können wir selbst in diese Entwicklung eingreifen?
ECC M1	7./8. Klasse	Drei Schulstunden (135 Minuten)	Wir gestalten unsere Welt aktiv durch gute Software: Einbindung aktueller Themen wie Umweltschutz, Gesundheit, Stärkung von Minderheiten etc.

Tab. 1: Das Konzept des EmiLe Computer Clubs über die Klassenstufen drei bis acht

In allen Kursen wird haptisches Material in Form von Lernkarten verwendet, welches den SuS erlaubt, selbstständig und im eigenen Tempo zu arbeiten.[4] Die Lehrkraft erläutert und zeigt zu Beginn der Unterrichtsstunde lediglich das Thema, die verwendeten Materialien sowie das Ziel der jeweiligen Einheit.[5] Im weiteren Verlauf der Stunde wird die Lehrkraft nur im Bedarfsfall hinzugerufen, wenn die SuS (z.B. aufgrund eines Programmierfehlers) selbst nicht mehr weiterkommen. Damit unterstützt das Kurskonzept das Montessori-Prinzip „Hilf mir, es selbst zu tun.“.[6]

3 Ausgewählte Beispiele

Im Folgenden werden einige ausgewählte Beispiele für die Anwendung von Material im ECC vorgestellt, mit der die Autorin gute Erfahrungen gemacht hat.[7] In jedem Fall ist es absolut empfehlenswert, die Werkzeuge und Lernkarten vorher selbst auszuprobieren. So kann auf Schwierigkeiten, auf welche die SuS stoßen werden, im laufenden Unterricht souverän reagiert werden.

3.1 ECC P1 (3./4. Klasse)

In dieser Klassenstufe wird der ECC mit Ipads in den Klassenräumen der Grundschule durchgeführt. Die Nutzung der Ipads hat mehrere Vorteile – so müssen die SuS kein personalisiertes Kennwort eingeben (was bei Windows-PCs der Fall wäre), und die

[4] Auch das Ergebnis der Arbeit kann durch die SuS unmittelbar selbst kontrolliert werden, was ein weiteres Qualitätsmerkmal im Sinne der Montessori-Pädagogik umsetzt [Sc16].

[5] Damit nimmt die Lehrkraft die Rolle einer „Begleitung“ ein, welche die SuS auf dem Weg des Wissenserwerbs unterstützt [Ei08], [Sc16].

[6] Die Montessori-Pädagogik zeichnet sich durch deutlich weitreichendere Prinzipien aus, auf die jedoch in diesem Artikel nicht eingegangen werden soll. Details können z.B. in [Sc16] nachgelesen werden.

[7] Die vorgestellten Beispiele dürfen nicht per se als montessorische Freiarbeitsmaterialien im Sinne von Maria Montessoris Pädagogik verstanden werden – eine solche Einordnung würde einer dezidierteren wissenschaftlichen Untersuchung und Diskussion bedürfen.

Bedienung erfolgt per „Fingerbedienung“ direkt auf dem Display. So müssen sich die SuS nicht an die Bedienung einer Maus am PC gewöhnen und können sich ganz auf die Inhalte konzentrieren.

Die ersten Stunden werden in der Regel zunächst mit **analogem Programmieren** gestaltet.[8] Dies kann, je nach Wetter, im Gebäude oder im Pausenhof geschehen.

Im zweiten Schritt hat es sich bewährt, zunächst „angeleitet“ programmieren zu lassen. Dafür bietet beispielsweise **code.org**[9] einige passende Projekte („Hour of Code“) an. Das Eingeben langer Webseiten-Adressen mit dem Ipad kann für die SuS herausfordernd sein – daher werden die Webseiten mittels eines QR-Codes aufgerufen.[10] Dieser QR-Code wird ausgedruckt ausgelegt, die SuS können ihn über die Kamerafunktion des Ipads einlesen und werden direkt auf die entsprechende Webseite weitergeleitet.

Nach einigen Stunden kommt dann die sehr beliebte Programmierumgebung **Scratch**[11] zum Einsatz. Zunächst wird den SuS die Umgebung auf dem Beamer erläutert. Anschließend arbeiten die SuS im eigenen Tempo mit Lernkarten, z.B. direkt von der Scratch-Foundation.[12] Hier lässt sich bereits die recht unterschiedliche Arbeitsweise der SuS erkennen – während einige die Lernkarten dankend annehmen und sich recht strikt an die Anleitungen halten, erarbeiten andere SuS von Anfang an eigene Ideen und experimentieren mit ihnen unbekannten Codeblöcken von Scratch. Am Ende der Stunde können die SuS die erstellten Spiele auf Wunsch veröffentlichen, so dass sie im Internet erreichbar sind. Schließlich werden die Spiele untereinander ausprobiert und online „likes“ vergeben.

3.2 ECC P2 (5./6. Klasse)

In dieser Klassenstufe wird der Unterricht in den Computerraum auf stationäre PCs verlegt. Ein bei den SuS sehr beliebtes Werkzeug ist die **Minecraft Education Edition**[13], die lokal auf den Schulrechnern installiert wird. In dieser, speziell für die Lehre entwickelten Minecraft-Version sind viele vorbereitete Lektionen bereits enthalten. Für den Einstieg bietet sich z.B. das Projekt „Stunde des Codes 2020: Inklusion“ an – dieses ist auch ohne Login in der Demo-Version spielbar.[14] Durch die Nutzung eines programmierbaren Helfers (dem sogenannten „Agenten“) werden dabei verschiedene kleine Rätsel in visueller, blockbasierter Programmierung gelöst, und so zwei verfeindete

[8] Hilfreiche Handreichungen sind beispielsweise unter https://www.csunplugged.org/de/ oder https://tueftelakademie.de/fuer-lehrende/fortbildungen/analoges-programmieren-was-ist-code/ verfügbar.
[9] https://code.org/hourofcode/overview
[10] Für jede Webseite kann mittels kostenfreier Online-Tools (wie z.B. dem qrcodemonkey https://www.qrcode-monkey.com/de/) oder bereits im Browser enthaltener Tools ein QR-Code erstellt werden.
[11] https://scratch.mit.edu
[12] https://resources.scratch.mit.edu/www/cards/de/scratch-cards-all.pdf
[13] https://education.minecraft.net
[14] Die in der Demo-Version spielbaren Projekte variieren von Jahr zu Jahr.

Dörfer versöhnt. Ebenfalls sehr empfehlenswert für den Einstieg ist die Lektion „Minecraft Programmiererstadt“, für die allerdings ein Login nötig ist.[15]

Ebenfalls in dieser Klassenstufe zum Einsatz kommt der **Calliope Mini**[16], ein programmierbarer Microcomputer. Im ECC wird dabei der Editor des Fraunhofer-Institutes Open Roberta[17] genutzt. Auch hier wird den SuS zunächst der Calliope Mini sowie die Programmierumgebung erläutert, anschließend bearbeiten die SuS ihre Projekte mittels Lernkarten[18] im eigenen Tempo. Im späteren Verlauf können die SuS eigene Projektideen entwickeln und umsetzen.

3.3 ECC M1 (7./8. Klasse)

In dieser Klassenstufe wird es deutlich anspruchsvoller – trotzdem wird weiterhin blockbasiert programmiert, was auch QuereinsteigerInnen einen leichten Einstieg garantiert.

So wird den SuS mittels des Tools **Machine Learning for Kids**[19] in verschiedenen Projekten spielerisch die Wirkweise von Künstlicher Intelligenz (KI) vermittelt. Auch hier wird im eigenen Tempo mit Lernkarten[20] gearbeitet, die Machine Learning (ML)-Modelle werden dabei mit der bereits bekannten Programmierumgebung Scratch verknüpft. Die Unterrichtseinheiten werden jeweils mit kurzen Videos über KI und ML ergänzt. So lernen die SuS spielerisch, was überwachtes und unüberwachtes Lernen bedeutet, wie eine KI trainiert wird, wodurch ein „Bias“ entstehen kann und welche Folgen dieser hat.

Ebenfalls zum Einsatz kommt das Werkzeug **App Inventor**[21]. Mit diesem Werkzeug können die SuS eigene Apps für das Smartphone oder Tablet erstellen. Hier wird ebenfalls zunächst mit Lernkarten von AppCamps gearbeitet, einige ausrangierte Android-Tablets und -Handys dienen dazu, die Apps direkt zu testen.[22] Im späteren Verlauf können die SuS eigene Projektideen, z.B. aus dem Bereich Umwelt-, Gesundheitsschutz oder Menschenrechte, entwerfen und umsetzen.

---

[15] Der Login ist an eine kostenpflichtige Lizenz gebunden und kann über die Bildungseinrichtung beantragt werden.
[16] https://calliope.cc
[17] https://lab.open-Roberta.org
[18] Z.B. unter https://calliope.cc oder auch https://appcamps.de
[19] https://machinelearningforkids.co.uk
[20] https://machinelearningforkids.co.uk/#!/worksheets
[21] Über die Webseite http://code.appinventor.mit.edu ist der App Inventor für SuS auch ohne Google-Account nutzbar.
[22] Wahlweise können die SuS ihre Apps auch direkt auf den eigenen, mitgebrachten Mobilgeräten ausprobieren.

4 Zusammenfassung und Ausblick

Ein entscheidender Punkt an der Montessori-Pädagogik ist die Arbeit mit „begreifendem Material“. Jedoch war damals, als Maria Montessori ihre Pädagogik entwickelte, virtuelles Material im Sinne von Software nicht vorgesehen.[23] Heute dagegen sind Computer in der Lehre nicht mehr wegzudenken.

Im Falle des hier vorgestellten Ansatzes des ECC steht das begreifende Material im überwiegenden Teil als virtuelles Material (im Sinne der Software, in der sich die SuS ausprobieren können) zur Verfügung. Frühere Studien geben Hinweise darauf, dass es für die Lernenden keinen Unterschied macht, ob Material physisch oder virtuell ausgeprägt ist [KTW07]. Dies lässt sich auch im ECC beobachten – dort funktioniert das Programmieren lernen mittels virtueller Materialien sehr gut. Durchweg lassen sich eine sehr hohe Eigenmotivation sowie große Lernerfolge der SuS beobachten.

Zum Ende des Schuljahres 2021/22 wurden die Kurse des ECC dreimal (zu je einem Halbjahr) durchlaufen und erfreuen sich bei den SuS großer Beliebtheit.[24] Bei der Auswahl der SuS für die Kurse wird ein besonderer Fokus auf Mädchen gelegt, und auch die Autorin selbst dient als Informatik-Lehrkraft den Mädchen als positives weibliches Vorbild.[25] Dadurch soll das Interesse für Informatik bei Mädchen gezielt gefördert werden, was langfristig der Erhöhung des Frauenanteils im MINT-Bereich zugutekommen soll. Eine wissenschaftliche Begleitung des Projektes erfolgt bisher nicht, wäre aus Sicht der Autorin jedoch wünschenswert. So könnten beispielsweise mögliche positive Auswirkungen des ECC auf das Interesse und die Fähigkeiten der TeilnehmerInnen für die Informatik über die Jahre langfristig untersucht werden.

Literaturverzeichnis

[AG14] Armoni, M.; Gal-Ezer, J.: Early computing education - Why? What? When? Who? In: ACM Inroads Bd. 5 Nr. 4, S. 54–59, 2014.

[Ei08] Eichelberger, H.: Handbuch zur Montessori-Didaktik. 4. Auflage, StudienVerlag, Innsbruck, 2008.

[Em22] EmiLe Montessori-Schulverein München-Südost e.V.: EmiLe Schulprofil, https://www.emile-montessori.de/paedagogische-grundlagen/schulprofil/, Stand: 24.06.2022.

[Ge22] Gesellschaft für Informatik (GI): GI bezieht Stellung zur ergänzenden Empfehlung zur KMK-Strategie „Bildung in der digitalen Welt“, https://gi.de/meldung/kmk-strategie, Stand:

[23] Maria Montessori verstarb 1952, der erste PC wurde mit dem Apple I im Jahr 1976 vorgestellt.

[24] Die Anzahl der Interessenten für die Kurse ist regelmäßig deutlich höher als die verfügbaren Plätze.

[25] Insbesondere ab der P2 kann beobachtet werden, dass sich mehr Jungen als Mädchen in die Kurse des ECC einwählen. Ein bestimmter Anteil der Kursplätze wird daher speziell für Mädchen vorgehalten und diese werden aktiv angesprochen, ob sie den Kurs wählen möchten. Werden die vorgehaltenen Plätze durch Mädchen nicht ausgeschöpft, werden auch diese Kursplätze an Jungen vergeben.

26.06.2022.

[He15] Heikki, T.: Gender Imbalance in Computing: Lessons from a Summer Computer Camp. In: ACM Inroads Bd. 6, Nr. 4, S. 22–23, 2015.

[HG22] Hubwieser, P.; Geldreich, K.: Algorithmen für Kinder (AlgoKids), Abschlussbericht eines Kooperationsprojektes. Technische Universität München, 2022.

[KTW07] Klahr, D.; Triona, L.; Williams, C.: Hands on What? The Relative Effectiveness of Physical Versus Virtual Materials in an Engineering Design Project by Middle School Children. In: Journal of Research in Science Teaching Bd. 44, Nr. 1, S. 183–203, 2007.

[Ku21] Kultusministerkonferenz: Lehren und Lernen in der digitalen Welt: Die ergänzende Empfehlung zur Strategie „Bildung in der digitalen Welt“, 2021.

[Mu22] Münchner Lehrer- und Lehrerinnenverband (MLLV): Neues Fach Informatik als Herausforderung für Lehrer/innen und Schüler/innen, https://mllv.bllv.de/chronik/2020/2020-02-maerz-april/neues-fach-informatik-als-herausforderung-fuer-lehrerinnen-und-schuelerinnen, Stand: 25.06.2022.

[SHF21] Schwarz, R.; Hellmig, L.; Friedrich, S.: Informatikunterricht in Deutschland – eine Übersicht. In: Informatik Spektrum Bd. 44, Nr. 2, S. 95–103, 2021.

[Sc16] Schumacher, E.: Montessori-Pädagogik verstehen, anwenden und erleben – Eine Einführung. Beltz Verlag, Basel u.a., 2016.

[SC22] Sun, C.; Clarke-Midura, J.: Recruiting K-12 Youth into Computer Science: Insights from a Multi-Year Research Project. In: ACM Inroads Bd. 13, Nr. 2, S. 22–29, 2022.

[St22] Staatsinstitut für Schulqualität und Bildungsforschung München: Fachlehrpläne Informatik an der Mittelschule Bayern, https://www.lehrplanplus.bayern.de/schulart/mittelschule/fach/informatik/inhalt/fachlehrplaene, Stand: 25.06.2022.

GI Startup Platform – Community Get Together

GI Startup Platform – Community Get Together

Wolfram Wingerath[1] Viktor Leis[2] Felix Gessert[3]

Abstract: The GI Startup Platform is a GI Junior Fellow initiative that addresses founders and startups in the field of computer science and pursues two specific goals. First, we would like to develop a central point of contact for information on all phases of a company's lifecycle, from founding over capital search & funding to exit & sale. Secondly, we want to connect stakeholders from all areas and bring founders together with other founders, mentors, investors, potential customers, and students looking for a job or internship. In our first Community Get Together, we invite everybody interested in startups to share their story, look for potential partners, engage in discussions, or simply enjoy the free snacks and listen. Beyond a brief introductory presentation, we will have several invited talks from community members with generous networking breaks in between to spark dialogue among the participants beyond the standard after-talk Q&A.

Additional info available at: https://startup.gi.de

1 GI Startup Platform: Mission Statement & Workshop Goal

In Germany, despite a rapidly growing tech startup scene, founding companies out of universities is still uncommon. Especially in the computer science departments of universities, the much-needed talent for founding future high-tech firms does exist, but at the same time there is enormous uncertainty among these potential founders about the right approach as well as a lack of network and business know-how.

Our initiative „GI Startup Platform" is aimed at people interested in founding, founders and startups in the field of computer science and pursues two specific goals. Firstly, we would like to develop a central point of contact for information on all entrepreneurial phases from founding to capital search & funding to exit & sale. The information offered includes helpful web portals and online courses as well as book recommendations, templates for contracts and field reports. Secondly, we want to connect stakeholders from all areas and bring together founders with other founders, mentors, investors, potential customers, and students looking for a job or internship.

With our first workshop, we aim to connect practitioners as well as researchers within and beyond the German Informatics Society (Gesellschaft für Informatik, GI).

[1] Universität Oldenburg, Department für Informatik (Data Science), Escherweg 2, 26121 Oldenburg, Germany, wolle@uol.de
[2] Friedrich-Alexander-Universität Erlangen-Nürnberg, Department Informatik (Datenmanagement), Martensstraße 3, 91058 Erlangen, Germany, viktor.leis@fau.de
[3] Baqend GmbH, Stresemannstraße 23, 22769 Hamburg, Germany, felix.gessert@baqend.com

doi:10.18420/inf2022_105

Trustworthy AI in Science and Society

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1227

The Influence of Training Parameters on Neural Networks' Vulnerability to Membership Inference Attacks

Oussama Bouanani[1], Franziska Boenisch[2]

Abstract: With Machine Learning (ML) models being increasingly applied in sensitive domains, the related privacy concerns are rising. Neural networks (NN) are vulnerable to, so-called, membership inference attacks (MIA) which aim at determining whether a particular data sample was used for training the model. The factors that render NNs prone to this privacy attack are not yet fully understood. However, previous work suggests that the setup of the models and the training process might impact a model's risk to MIAs. To investigate these factors more in detail, we set out to experimentally evaluate the influence of the training choices in NNs on the models' vulnerability. Our analyses highlight that the batch size, the activation function, and the application and placement of batch normalization and dropout have the highest impact on the success of MIAs. Additionally, we applied statistical analyses to the experiment results and found a highly positive correlation between a model's ability to resist MIAs and its generalization capacity. We also defined a metric to measure the difference in the distributions of loss values between member and non-member data samples and observed that models scoring higher values on that metric were consistently more exposed to the attack. The latter observation was further confirmed by manually generating predictions for member and non-member samples producing loss values within specific distributions and launching MIAs on them.

Keywords: privacy; machine learning; neural networks; membership inference

1 Introduction

With machine learning (ML) being increasingly applied in sensitive domains, such as health care [PC19], protecting privacy of the models gains importance. So far, it has been shown that ML models are prone to leak private information about their training data [FJR15, Ga18, Sh17]. Membership Inference Attacks (MIA) [Sh17] manifest in the ability of an adversary to determine whether a particular data sample has been used to train a target model. Therefore, the adversary solely requires access to the target model and the data sample under test. There exist two broad access scenarios for MIAs. In the more common black-box setting, the adversary only has access to the model's output, i.e. its predictions on the data sample. In white-box settings, the adversary has full access to the model and its parameters, which can be used to craft more successful attacks.

Previous work investigated factors influencing the success of MIAs. It was observed that overfitting puts a model at more risk [Sh17, Ye18], and that outlier data samples were more

[1] Freie Universität Berlin, Germany oussama.bouanani@fu-berlin.de
[2] Fraunhofer AISEC, Department of Secure Systems Engineering, Germany franziska.boenisch@aisec.fraunhofer.de

doi:10.18420/inf2022_106

vulnerable even in well-generalized models [Lo18]. Additionally, different loss values in predictions on training and test data samples were highlighted as an important factor that the attack can exploit [Sa19a, Sh17, Ye18].

While there exist methods that explicitly aim at reducing a model's vulnerability to MIAs, such as Differential Privacy [Dw06] or regularization techniques [NSH18, Sh17, Sr14], the role of architectural choices and training parameters in protecting or exposing a model has not yet been extensively studied. Therefore, we set out to investigate the influence of individual training choices, including hyper-parameters, layer-wise model configurations, optimizers, and training data transformations on the vulnerability of NN models to MIAs. We, therefore, analyzed the impact of the training configurations by tracking different metrics throughout the experiments, including a metric we defined to measure the difference in the distributions of loss values between member and non-member samples. Additionally, to confirm the important role of loss values as highlighted in previous work [Sa19a, Sh17, Ye18], we manually generated predictions for data samples. This allowed us to determine the resulting loss value and to study the attack success solely based on their distributions and independent of additional factors that might have an influence on the target model's vulnerability.

Our results suggest that smaller batch sizes contribute in lowering NN models' exposure to MIAs and that using *tanh* as an activation function significantly increases the attack's success. While weight regularization did not lead to a remarkable impact on the model's vulnerability to MIAs, dropout had a varied influence depending on which layers of the NN it was applied to. As a general trend, our analyses show that the training configurations that result in decreased generalization capacities of a model allowed the MIAs to gain more success. We also observed that the architectural choices that tend to increase the loss distribution difference between member and non-member samples consistently rendered the model more vulnerable. The latter observation was further confirmed with the manually generated probabilities experiment. As a consequence, we found that incorrectly classified samples are significantly more prone to membership inference than correctly classified ones and that samples whose predictions produced the lowest loss values were more resilient to the attack.

The scope of this work is limited to studying MIA vulnerability in NN classifiers, and only architectural choices that would allow models to maintain a good performance were considered.

2 Background and Related Work

MIAs were first introduced by Shokri et al. [Sh17]. To predict a data sample's membership, Shokri proposed using a binary classifier that can distinguish between member and non-member data samples. For training this binary classifier, so-called shadow models that imitate the behavior of the target model but for which the used training data samples are

known, were created. The shadow training technique was adopted and further developed in subsequent studies [Tr19, Sa19b].

In previous work, some factors that influence the success of MIAs have already been investigated. The role of overfitting was highlighted as one of the aspects that weakens a model against MIAs [Sh17, Ye18]. Additionally, it was observed that outlier data samples were more prone to the attack even in well-generalized ML models [Lo18] and that MIAs found more success when targeting samples with a more considerable influence on the decision boundary of a classification model [Tr19]. The general properties of the dataset used to train a ML model, including the amount of classes and in-class standard deviation, were highlighted as additional factors that influence a model's vulnerability to MIAs [Tr19].

Shokri et al. [Sh17] observed a significant difference in the distributions of the outputs produced between member and non-member instances and marked it as one of the aspects that their attack was exploiting. In a later work, Sablayrolles et al. [Sa19a] presented a theoretical proof that the optimal MIA depends only on its ability to set a decisive threshold that separates loss values of member and non-member sample predictions.

As mitigation of MIAs, regularization techniques were suggested, including adversarial regularization [NSH18], L2-regularization [Sh17], and dropout [Sr14]. However, while these techniques were able to reduce a model's exposure to this attack, they also contributed to lowering the model's utility by reducing its classification accuracy and, therefore, often failing to provide a reasonable privacy-utility trade-off [SH20].

3 Experiments

To evaluate the impact of architectural choices and training parameters on MIAs, we experimentally trained different NN models, launched MIAs on them, and measured the models' vulnerability.

3.1 Metrics

We tracked three different metrics throughout the experiments. For a target model, we define $train_acc$, $test_acc$ as the training accuracy and testing accuracy of the model respectively.

Attacker advantage (attack adv.) Inspired by definition 4. in [Ye18] and bounded within $[0, 1]$, the attack adv. is defined as the maximum absolute difference of the true positive rate (TPR) and false positive rate (FPR) ($max(|TPR - FPR|)$) that the attack model can achieve over all possible decision thresholds to separate member from non-member data samples. This metric allows us to consider the attack model in a state where it is able to find the optimal theoretical decision threshold that yields the highest correct membership predictions.

Train-test accuracy gap (acc. gap) Tracks the gap in the prediction performance of the target model between member (training data where the training accuracy is acc_{train}) and non-member (test data where the testing accuracy is acc_{test}) samples. Defined as $|acc_{train} - acc_{test}|$, it highlights the degree of overfitting of a target model to its training data.

Train-test loss distribution difference (loss diff.) To quantify the difference in the loss distributions between member and non-member samples, we applied the *Kruskal-Wallis* H-test [Do08a]. This non-parametric method does not assume any specific underlying distribution in the different groups. Its results express to what degree the loss values among members and non-members differ: the higher the outcome, the larger the discrepancies between the two distributions.

In order to understand the connection between the performance of MIAs and the prediction quality of the target model, we looked into the correlations between the attack adv. and the two other aforementioned metrics. The correlations were determined using the *Spearman's rank correlation coefficient* [Do08b].

3.2 Experimental Setup

The experiments were conducted with TensorFlow (TF) 2.4.1 [Te15] using a baseline multi-class classifier convolution NN [CR93] inspired by the VGG-Net [SZ15] architecture. It consists of three consecutive convolution-blocks where each block has two convolution layers followed by a max pooling layer where convolution layers have the same kernel size of 3x3 and an increasing depth within each convolution-block. The convolutions are then followed by two fully-connected layers. The model was trained on the CIFAR-10 [Kr12] Dataset over 100 epochs with a batch size of 64, the layers have the weights initialized from the Glorot Uniform distribution [GB10] and use ReLU as the activation function. The baseline was trained using Stochastic Gradient Descent with momentum = 0.01 and learning rate= 0.01, and a categorical cross-entropy loss function.

In each experiment, one training parameter or architectural aspect of the baseline model was altered according to Table 1, while leaving the rest unchanged. We also experimented with different data sets, namely MNIST [De12], Fashion-MNIST [XRV17], SVHN [SCL12]. A threshold-based MIA, requiring no training, and four different trainable attack model types (logistic regression, random forest, multi-layer perceptron, k-nearest neighbors) from the TF Privacy library [Te19], were then launched on each resulting trained variation of the model. The threshold-based attack takes the prediction vector of a data point x from a target model

Aspect	**Variations**
Training Epochs	15, 30, 50, 100, 150, and 200.
Batch Size	32, 48, 64, 80, 96, 128, 144, 256, and 512.
Data Augmentation [TN18] [Le19]	None, Translation, Zoom, Rotation, Shear, Inversion, Brightness, Contrast, and Saturation.
Activation Function	ReLU [ESH19], leaky-ReLU [Li19], sigmoid [BK19], and tanh [LDH19].
Weight Initialization	Glorot Uniform [GB10], Glorot Normal [GB10], He Normal [He15], He Uniform [He15], Lecun Normal [Le98], and Lecun Uniform [Le98].
Weight Regularization	L1 [SFR07], L2 [vL17], and L1L2 with $\lambda \in \{l = 0.01, m = 0.005, s = 0.0025\}$.
Batch Normalization [IS15] with momentum m [Yo20]	Pre: Applied before activations with $m \in \{l = 0.99, m = 0.5, s = 0.1\}$. Post: Applied after activations with $m \in \{l = 0.99, m = 0.5, s = 0.1\}$.
Dropout [Sr14] with threshold t	Conv: Applied after each convolution layer with $t \in \{l = 0.3, m = 0.2, s = 0.1, vs = 0.05\}$. Block: Applied after each two convolution layers and max pooling layer with $t \in \{l = 0.3, m = 0.2, s = 0.1, vs = 0.05\}$. Dense: Applied once after dense layers with $t \in \{l = 0.5, m = 0.25, s = 0.1\}$.
Optimization Algorithm with learning rate lr, momentum m	$lr \in \{0.1, 0.05, 0.02\}$: Stochastic Gradient Descent (SGD) [Ru16], and SGD with Nesterov (SGDn) [Su13]. $lr \in \{0.01, 0.005, 0.001\}$: SGD with $m = 0.9$ (SGDm), and SGD with both Nesterov and $m = 0.9$ (SGDmn). $lr \in \{0.01, 0.005, 0.001\}$: RMSProp, Adam [KB15], Adamax (variant of Adam), Nadam (variant of Adam with Nesterov).
Loss Function	Categorical Cross Entropy, KL-Divergence, and Poisson Loss.

Tab. 1: Training parameters and architectural choices as variations of the baseline model in the experiments.

as input, calculates the loss value of the prediction and finally compares it to a threshold value to decide whether x was a member or non-member. The trained attacks use prediction vectors and loss values from both member and non-member sample data points for training to be able to predict the membership status of another data point.

In addition to studying the effect of concrete training parameters, we wanted to confirm the determining role of the loss values as highlighted in [Sa19a] independent of any training factors. Therefore, we set up two additional experiments where we manually generated prediction vectors for both member and non-member instances of the CIFAR-10 dataset. We specified the vectors such that they always predict the true class of each instance. This renders the acc. gap of the generated predictions to zero, also minimizing the role it plays in the MIAs' success.

In the first experiment, we started by assigning the same true class probability to member and non-member sample. For example, in a dataset of dimension 3 (possible classes are 1, 2 and 3), a prediction vector of a data point x of class 2 is $pred_x = [0, 1, 0]$, whether x is a member or non-member. Then, we successively reduced the true class probability for non-member samples until it was under 50%. So a non-member data point y of class 1 would progressively have its prediction vector changed from (originally) $pred_y = [1, 0, 0]$ to $pred_y = [0.95, 0.025, 0.025]$ then $pred_y = [0.9, 0.05, 0.05]$ until the difference reaches 50%, rendering the prediction vector of y to be $pred_y = [0.5, 0.25, 0.25]$. In the second experiment, we started with the same prediction vectors as in the first experiment. Afterwards, we randomly and progressively picked non-member instances and reduced their true class probability by 10%. In each combination of both experiments, we launch MIAs at regular intervals to assess the vulnerability presented by the prediction vectors and the resulting loss values.

3.3 Experimental Results

Notable Architectural Choices The baseline variations provided different results for MIAs, depending on the training choice handled in each experiment. Figure 1 shows the performance of the threshold-based MIA on variations of the data augmentation technique, weight regularization, batch size, activation function, batch normalization, and dropout experiments.

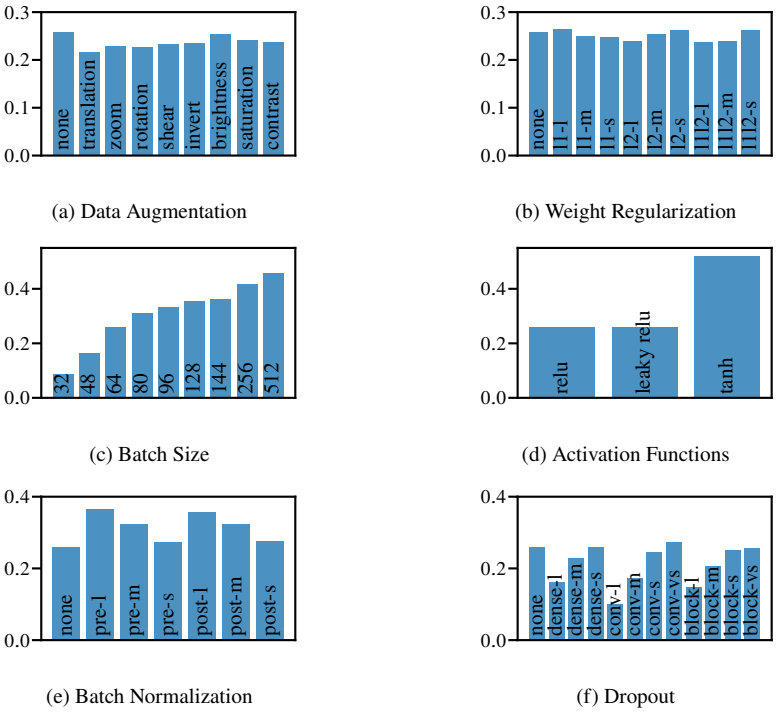
While applying different data augmentation techniques allowed the baseline model to lower its exposure to MIAs in various degrees, only *translation* was consistently providing the highest protection against all attack model types. Increasing the batch size value consistently increases the success of the MIAs while increasing the training epochs up to 100 increased the threat of attacks which was diminished when training for up to 150 or 200 epochs. Regarding activation functions, no considerable difference in the MIAs performance was observed when using either ReLU or leaky ReLU while a significant increase in the attacks' success was observed when using tanh. Additionally, the learning rate of the optimizer that

provides the best protection depended heavily on the optimization algorithm. The Adamax optimizer with a learning rate of 0.0001 was the highest protective implementation.

Applying weight regularization didn't have a significant impact on the vulnerability of the baseline model as all attack models were able to perform similarly. Adding batch normalization to the baseline's architecture allowed to improve its performance, however it rendered it more vulnerable to the MIAs. Dropout had a varied impact on the success of the attacks, depending on where it is implemented and the threshold value. For instance, applying dropout after the dense layers of the baseline provided no significant protection for threshold values lower than 0.5 while applying it after convolution layers, with small threshold values (< 0.1) allowed the baseline model to decrease its exposure to MIAs.

We provide results for all experiments by reporting the tracked metrics for each varied hyper-parameter in the Appendix section.

Consistent Tendencies The correlation analysis, described in subsection 3.1, allowed to gain insights onto why specific implementations provided more protection than others.



(a) Data Augmentation

(b) Weight Regularization

(c) Batch Size

(d) Activation Functions

(e) Batch Normalization

(f) Dropout

Fig. 1: Threshold MIA attack adv. (y-axis) on experiment variations (x-axis). See Table 1 for variations.

As illustrated in Figure 2, the most consistent tendency was a strong positive correlation between the attack adv. and the loss diff. This shows that MIAs achieved more success when loss values present a more significant difference between member and non-member instances.

The acc. gap correlation with the attack adv. was not that consistent when looking at the attack results on the entire dataset. However, when considered class-wise, we observed a strong positive correlation. This observation highlights that a higher gap in the performance between member and non-member instances in specific classes exposes the instances of these classes more to the attacks.

Furthermore, the MIAs were much more successful at inferring the membership status of the misclassified instances than correctly classified ones. Similar tendencies have been observed by [RL21]. When considering the performance of MIAs over data grouped by their loss percentiles, the highest attack adv. values were spotted in the instances with the highest loss values. An example of this behavior is depicted in Figure 3 for the batch size experiment variations.

Manually Generated Probabilities Figure 4 illustrates the performance of the MIA on the manually generated variations described in subsection 3.2. Results show that the MIAs achieve a zero attack adv. when the predictions of member and non-member instances are on the same distribution. However, even with smallest probability difference between member and non-member samples, the MIAs have achieved full attack adv. as they become able to set a decisive threshold that separates member and non-member probabilities and their resulting loss values. Additionally, the success of MIAs aligns with the percentage of non-member samples having distinct predictions to the member samples. These results

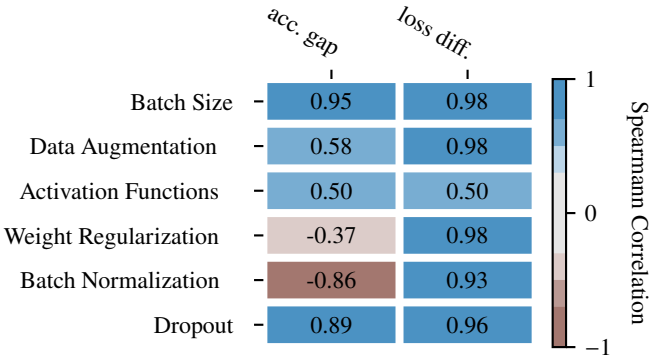


Fig. 2: Correlation between threshold MIA attack adv., and acc. gap or loss diff. for various experiments on the whole dataset.

confirm the observations in past work [Sa19a, Sh17, Ye18] regarding the important role that the loss has in the success of MIAs. They also extend the results by providing the insight that this influence is not due to other factors, such as the training configuration or the overall performance of the target model.

4 Conclusions

The use of sensitive information to train ML models exposes them to privacy-concerns. Understanding the factors that influence a model's resilience against privacy-breaching attacks is essential to protect them from such risks. In the experiments, we investigated the role of a NN's training choices on its vulnerability to MIAs. Our results highlight that using high batch size values during training increases the model's exposure to this privacy attack while having ReLU or leaky ReLU as the activation function provides more protection than using tanh. In regards to regularization techniques, we were able to show that implementing batch normalization exposes NNs to more MIA risk. Concerning dropout, our experiments

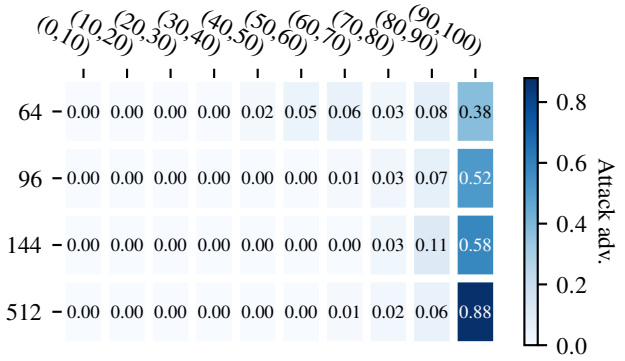


Fig. 3: Threshold MIA attack adv. when launched on data grouped by loss percentile in batch size experiments.



Fig. 4: Threshold MIA attack adv. (y-axis) on manually generated probability variations (x-axis): *Left*: True class probability difference. *Right*: Different non-member prediction percentage.

showed that it does allow NNs to gain significant protection against the attack, but only when it is enabled after convolution layers.

The results of the experiments show that the success of MIAs is strongly correlated with low generalization of a model's predictions, resulting from the use of specific training configurations. We were able to confirm the important role of the loss with the attack models achieving maximum attack adv. when launched on manually generated loss values that differ between member and non-member instances. Data samples that are most prone to MIAs, therefore, usually are the ones that are incorrectly classified, have a relatively high loss, and stem from classes on which the target model exhibits a low-generalization capacity.

Bibliography

[BK19] Bawa, Vivek Singh; Kumar, Vinay: Linearized sigmoidal activation: A novel activation function with tractable non-linear characteristics to boost representation capability. Expert Systems with Applications, 120:346–356, 2019.

[CR93] Chua, L.O.; Roska, T.: The CNN paradigm. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 40(3):147–156, 1993.

[De12] Deng, Li: The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. IEEE Signal Processing Magazine, 29(6):141–142, 2012.

[Do08a] Dodge, Yadolah: Kruskal-Wallis Test. In: The Concise Encyclopedia of Statistics, pp. 288–290. Springer New York, New York, NY, 2008.

[Do08b] Dodge, Yadolah: Spearman Rank Correlation Coefficient. In: The Concise Encyclopedia of Statistics, pp. 502–505. Springer New York, New York, NY, 2008.

[Dw06] Dwork, Cynthia: Differential Privacy. In (Bugliesi, Michele; Preneel, Bart; Sassone, Vladimiro; Wegener, Ingo, eds): Automata, Languages and Programming. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–12, 2006.

[ESH19] Eckle, Konstantin; Schmidt-Hieber, Johannes: A comparison of deep networks with ReLU activation function and linear spline-type methods. Neural Networks, 110:232–242, 2019.

[FJR15] Fredrikson, Matt; Jha, Somesh; Ristenpart, Thomas: Model Inversion Attacks That Exploit Confidence Information and Basic Countermeasures. In: Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security. CCS '15, Association for Computing Machinery, New York, NY, USA, p. 1322–1333, 2015.

[Ga18] Ganju, Karan; Wang, Qi; Yang, Wei; Gunter, Carl A.; Borisov, Nikita: Property Inference Attacks on Fully Connected Neural Networks Using Permutation Invariant Representations. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. CCS '18, Association for Computing Machinery, New York, NY, USA, p. 619–633, 2018.

[GB10] Glorot, Xavier; Bengio, Yoshua: Understanding the difficulty of training deep feedforward neural networks. In (Teh, Yee Whye; Titterington, Mike, eds): Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. volume 9 of Proceedings of Machine Learning Research, PMLR, Chia Laguna Resort, Sardinia, Italy, pp. 249–256, 13–15 May 2010.

[He15] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian: Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, Santiago, Chile, pp. 1026–1034, 2015.

[IS15] Ioffe, Sergey; Szegedy, Christian: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37. ICML'15, JMLR.org, Lille, France, p. 448–456, 2015.

[KB15] Kingma, Diederik P.; Ba, Jimmy: Adam: A Method for Stochastic Optimization. In (Bengio, Yoshua; LeCun, Yann, eds): Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015. ICLR 2015, San Diego, CA, USA, p. X, 2015.

[Kr12] Krizhevsky, Alex: , Learning Multiple Layers of Features from Tiny Images, May 2012.

[LDH19] Lohani, Harshit Kumar; Dhanalakshmi, S.; Hemalatha, V.: Performance Analysis of Extreme Learning Machine Variants with Varying Intermediate Nodes and Different Activation Functions. In (Mallick, Pradeep Kumar; Balas, Valentina Emilia; Bhoi, Akash Kumar; Zobaa, Ahmed F., eds): Cognitive Informatics and Soft Computing. Springer Singapore, Singapore, pp. 613–623, 2019.

[Le98] LeCun, Yann; Bottou, Léon; Orr, Genevieve B.; Müller, Klaus-Robert: Efficient BackProp. In: Neural Networks: Tricks of the Trade, This Book is an Outgrowth of a 1996 NIPS Workshop. Springer-Verlag, Berlin, Heidelberg, p. 9–50, 1998.

[Le19] Lei, Cheng; Hu, Benlin; Wang, Dong; Zhang, Shu; Chen, Zhenyu: A Preliminary Study on Data Augmentation of Deep Learning for Image Classification. In: Proceedings of the 11th Asia-Pacific Symposium on Internetware. Internetware '19, Association for Computing Machinery, New York, NY, USA, 2019.

[Li19] Liu, Yisi; Wang, Xiaojun; Wang, Lei; Liu, Dongliang: A modified leaky ReLU scheme (MLRS) for topology optimization with multiple materials. Applied Mathematics and Computation, 352:188–204, 2019.

[Lo18] Long, Yunhui; Bindschaedler, Vincent; Wang, Lei; Bu, Diyue; Wang, Xiaofeng; Tang, Haixu; Gunter, Carl A.; Chen, Kai: Understanding Membership Inferences on Well-Generalized Learning Models. CoRR, abs/1802.04889:X, 2018.

[NSH18] Nasr, Milad; Shokri, Reza; Houmansadr, Amir: Machine Learning with Membership Privacy Using Adversarial Regularization. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security. CCS '18, Association for Computing Machinery, New York, NY, USA, p. 634–646, 2018.

[PC19] Price, W. Nicholson; Cohen, I. Glenn: Privacy in the Age of Medical Big Data. Nature Medicine, 25(1):37–43, January 2019.

[RL21] Rezaei, Shahbaz; Liu, Xin: On the Difficulty of Membership Inference Attacks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Virtual Event, pp. 7892–7900, June 2021.

[Ru16] Ruder, Sebastian: An overview of gradient descent optimization algorithms. CoRR, abs/1609.04747:X, 2016.

[Sa19a] Sablayrolles, Alexandre; Douze, Matthijs; Schmid, Cordelia; Ollivier, Yann; Jegou, Herve: White-box vs Black-box: Bayes Optimal Strategies for Membership Inference. In (Chaudhuri, Kamalika; Salakhutdinov, Ruslan, eds): Proceedings of the 36th International Conference on Machine Learning. volume 97 of Proceedings of Machine Learning Research, PMLR, Long Beach, CA, USA, pp. 5558–5567, 09–15 Jun 2019.

[Sa19b] Salem, Ahmed; Zhang, Yang; Humbert, Mathias; Fritz, Mario; Backes, Michael: ML-Leaks: Model and Data Independent Membership Inference Attacks and Defenses on Machine Learning Models. In: Network and Distributed Systems Security Symposium 2019. Internet Society, San Diego, CA, USA, p. X, 2019.

[SCL12] Sermanet, Pierre; Chintala, Soumith; LeCun, Yann: Convolutional neural networks applied to house numbers digit classification. In: Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012. IEEE 2012, Tsukuba Science City, Japan, pp. 3288–3291, 2012.

[SFR07] Schmidt, Mark; Fung, Glenn; Rosales, Rómer: Fast Optimization Methods for L1 Regularization: A Comparative Study and Two New Approaches. In (Kok, Joost N.; Koronacki, Jacek; de Mantaras, Raomon Lopez; Matwin, Stan; Mladenič, Dunja; Skowron, Andrzej, eds): Machine Learning: ECML 2007. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 286–297, 2007.

[Sh17] Shokri, Reza; Stronati, Marco; Song, Congzheng; Shmatikov, Vitaly: Membership Inference Attacks Against Machine Learning Models. In: 2017 IEEE Symposium on Security and Privacy (SP). IEEE Computer Society, San Jose, CA, USA, pp. 3–18, 2017.

[SH20] Shejwalkar, Virat; Houmansadr, Amir: Machine Learning with Membership Privacy via Knowledge Transfer. Privacy Preserving Machine Learning (PPML) 2020, 4:9, 2020.

[Sr14] Srivastava, Nitish; Hinton, Geoffrey; Krizhevsky, Alex; Sutskever, Ilya; Salakhutdinov, Ruslan: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15(56):1929–1958, 2014.

[Su13] Sutskever, Ilya; Martens, James; Dahl, George; Hinton, Geoffrey: On the Importance of Initialization and Momentum in Deep Learning. In: Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28. ICML'13, JMLR.org, Atlanta, GA, USA, p. III–1139–III–1147, 2013.

[SZ15] Simonyan, Karen; Zisserman, Andrew: Very Deep Convolutional Networks for Large-Scale Image Recognition. In (Bengio, Yoshua; LeCun, Yann, eds): Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015. ICLR, San Diego, CA, USA, p. X, 2015.

[Te15] TensorFlow: , TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. Software available from tensorflow.org.

[Te19] TensorFlow: , TensorFlow Privacy: Library for training machine learning models with privacy for training data, 2019.

[TN18] Taylor, Luke; Nitschke, Geoff: Improving Deep Learning with Generic Data Augmentation. In: 2018 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE Computer Society, Bengaluru, India, pp. 1542–1547, 2018.

[Tr19] Truex, Stacey; Liu, Ling; Gursoy, Mehmet Emre; Yu, Lei; Wei, Wenqi: Demystifying Membership Inference Attacks in Machine Learning as a Service. IEEE Transactions on Services Computing, 12:1–1, 2019.

[vL17] van Laarhoven, Twan: L2 Regularization versus Batch and Weight Normalization. CoRR, abs/1706.05350:X, 2017.

[XRV17] Xiao, Han; Rasul, Kashif; Vollgraf, Roland: Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. CoRR, abs/1708.07747:X, 2017.

[Ye18] Yeom, S.; Giacomelli, I.; Fredrikson, M.; Jha, S.: Privacy Risk in Machine Learning: Analyzing the Connection to Overfitting. In: 2018 IEEE 31st Computer Security Foundations Symposium (CSF). IEEE Computer Society, Oxford, UK, pp. 268–282, July 2018.

[Yo20] Yong, Hongwei; Huang, Jianqiang; Meng, Deyu; Hua, Xiansheng; Zhang, Lei: Momentum Batch Normalization for Deep Learning with Small Batch Size. In (Vedaldi, Andrea; Bischof, Horst; Brox, Thomas; Frahm, Jan-Michael, eds): Computer Vision – ECCV 2020. Springer International Publishing, Cham, Germany, pp. 224–240, 2020.

Appendices

A Experiment Results

The following tables include the tracked metrics for each varied hyper-parameter in each experiment. The attack adv. values are reported from the threshold attack.

A.1 Training Epochs

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
15	91.6%	70.8%	20.8%	1299.5	0.207
30	95.6%	71.4%	24.2%	1789.96	0.239
50	95.6%	70.4%	25.2%	2176.95	0.26
100	93.4%	71.7%	21.7%	2220.75	0.259
150	90.6%	70.7%	19.9%	1553.02	0.213
200	84.7%	64.2%	20.5%	1082.69	0.182

Tab. 2: Training Epochs experiment results.

A.2 Batch Size

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
32	69.1%	61.4%	7.7%	253.32	0.085
48	84.0%	65.2%	18.8%	840.16	0.163
64	93.4%	71.7%	21.7%	2220.75	0.259
80	97.3%	72.5%	24.8%	3693.78	0.311
96	98.4%	71.9%	26.5%	4534.32	0.332
128	99.3%	72.4%	26.9%	5293.78	0.354
144	99.4%	73.2%	26.2%	6060.22	0.36
256	100.0%	71.8%	28.2%	9022.63	0.418
512	100.0%	69.2%	30.8%	8033.68	0.457

Tab. 3: Batch Size experiment results.

A.3 Data Augmentation

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
None	93.4%	71.7%	21.7%	2220.75	0.259
Translation	91.6%	71.7%	19.9%	1524.19	0.216
Zoom	91.6%	70.2%	21.4%	1628.61	0.229
Rotation	91.9%	70.9%	21.0%	1625.93	0.227
Shear	93.1%	69.8%	23.3%	1711.83	0.234
Invert	91.4%	71.0%	20.4%	1746.08	0.236
Brightness	93.0%	69.6%	23.4%	1895.14	0.254
Saturation	92.9%	71.8%	21.1%	1994.48	0.241
Contrast	92.1%	69.2%	22.9%	1867.19	0.238

Tab. 4: Data Augmentation experiment results.

A.4 Activation Functions

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
ReLU	93.4%	71.7%	21.7%	2220.75	0.259
Leky ReLU	93.7%	71.5%	22.2%	2667.88	0.259
tanh	100.0%	74.7%	25.3%	4869.43	0.519

Tab. 5: Activation Functions experiment results.

A.5 Weight Initializers

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
Glorot Normal	93.6%	69.9%	23.7%	2429.06	0.27
Glorot Uniform	93.4%	71.7%	21.7%	2220.75	0.259
He Normal	92.0%	72.3%	19.7%	1538.81	0.219
He Uniform	92.2%	70.4%	21.8%	1669.12	0.236
Lecun Uormal	94.1%	71.0%	23.1%	1883.08	0.241
Lecun Uniform	94.1%	71.6%	22.5%	2312.83	0.261
Random Normal, $\sigma = 0.05$	91.7%	68.8%	22.9%	2316.04	0.255
Random Normal, $\sigma = 0.025$	94.0%	67.3%	26.7%	2954.44	0.308
Random Uniform, $y = 0.15$	88.6%	67.1%	21.5%	1604.32	0.233
Random Uniform, $y = 0.075$	91.7%	68.3%	23.4%	2172.48	0.26
Random Uniform, $y = 0.05$	93.9%	68.6%	25.3%	2599.92	0.273

Tab. 6: Weight Initializers experiment results.

A.6 Weight Regularizers

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
None	93.4%	71.7%	21.7%	2220.75	0.259
L1, $\lambda = 0.01$	93.7%	71.6%	22.1%	2334.11	0.265
L1, $\lambda = 0.005$	93.5%	69.2%	24.3%	2042.99	0.249
L1, $\lambda = 0.0025$	93.9%	70.3%	23.6%	2003.99	0.248
L2, $\lambda = 0.01$	93.7%	70.5%	23.2%	1813.37	0.239
L2, $\lambda = 0.005$	93.7%	70.9%	22.8%	2177.05	0.255
L2, $\lambda = 0.0025$	93.0%	71.2%	21.8%	2266.56	0.262
L1L2, $\lambda = 0.01$	94.4%	68.9%	25.5%	1763.01	0.237
L1L2, $\lambda = 0.005$	92.8%	70.7%	22.1%	1882.08	0.239
L1L2, $\lambda = 0.0025$	94.2%	69.8%	24.4%	2344.17	0.263

Tab. 7: Weight Regularizers experiment results.

A.7 Batch Normalization

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
None	93.4%	71.7%	21.7%	2220.75	0.259
Pre, $m = 0.99$	100.0%	81.6%	18.4%	2533.32	0.366
Pre, $m = 0.5$	100.0%	80.1%	19.9%	2404.66	0.324
Pre, $m = 0.1$	100.0%	78.0%	22.0%	2087.14	0.272
Post, $m = 0.99$	100.0%	80.9%	19.1%	2523.53	0.357
Post, $m = 0.5$	100.0%	79.4%	20.6%	2487.46	0.323
Post, $m = 0.1$	100.0%	77.4%	22.6%	2273.41	0.276

Tab. 8: Batch Normalization experiment results. Batch Normalization is applied with a different momentum value m before activations (Pre) and after activations (Post).

A.8 Loss Functions

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
Categorical CrossEntropy	93.4%	71.7%	21.7%	2220.75	0.259
KL Divergence	92.8%	70.2%	22.6%	1898.25	0.245
Poisson	100.0%	67.9%	32.1%	6893.19	0.441

Tab. 9: Loss Functions experiment results.

A.9 Dropout

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
$t = 0.3$	84.3%	71.7%	12.6%	773.64	0.16
$t = 0.2$	91.2%	72.1%	19.1%	1625.57	0.228
$t = 0.1$	93.4%	73.4%	20.0%	2151.14	0.26
$t = 0.05$	79.5%	77.6%	1.9%	266.71	0.099

Tab. 10: Dropout experiment results. Dropout is applied using threshold t after each convolution layer.

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
$t = 0.3$	86.8%	78.1%	8.7%	767.04	0.171
$t = 0.2$	92.0%	77.1%	14.9%	1543.35	0.244
$t = 0.1$	93.6%	74.5%	19.1%	2089.24	0.272
$t = 0.05$	86.3%	81.8%	4.5%	515.6	0.146

Tab. 11: Dropout experiment results. Dropout is applied using threshold t after each two consecutive convolution layers and max pooling layer.

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
$t = 0.5$	90.5%	81.2%	9.3%	1035.58	0.207
$t = 0.25$	94.3%	78.2%	16.1%	1715.06	0.25
$t = 0.1$	94.7%	76.1%	18.6%	1984.21	0.257

Tab. 12: Dropout experiment results. Dropout is applied once using threshold t after the dense layers.

A.10 Optimization Algorithms

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
SGD, $lr = 0.1$	93.4%	68.4%	25.0%	2220.75	0.274
SGD, $lr = 0.05$	99.1%	70.9%	28.2%	4759.34	0.346
SGD, $lr = 0.02$	100.0%	69.3%	30.7%	8084.81	0.439
SGDn, $lr = 0.1$	93.2%	68.2%	25.0%	2965.98	0.3
SGDn, $lr = 0.05$	99.6%	71.7%	27.9%	6598.82	0.385
SGDn, $lr = 0.02$	99.8%	67.7%	32.1%	6263.28	0.404
SGDm, $lr = 0.01$	93.4%	71.7%	21.7%	2220.75	0.259
SGDm, $lr = 0.005$	99.4%	72.1%	27.3%	5652.95	0.357
SGDm, $lr = 0.001$	99.8%	68.1%	31.7%	5814.34	0.412
SGDmn, $lr = 0.01$	93.9%	71.0%	22.9%	1693.64	0.225
SGDmn, $lr = 0.005$	99.4%	73.8%	25.6%	5164.7	0.339
SGDmn, $lr = 0.001$	99.8%	68.9%	30.9%	5246.01	0.396

Tab. 13: Stochastic Gradient Descent (SGD) Optimization Algorithm with learning rate lr experiment results, where SGDn is a variation with Nesterov [Su13] enabled, SGDm is a variation with momentum $m = 0.9$ enabled and SGDmn is a variation where both of the latter options are enabled.

Variation	Train Acc.	Test Acc.	Acc. Gap	Loss Diff.	Attack Adv.
RMSprop, lr = 0.0005	98.9%	78.3%	20.6%	6603.18	0.268
RMSprop, lr = 0.0002	99.4%	74.5%	24.9%	4057.4	0.297
RMSprop, lr = 0.0001	99.5%	72.7%	26.8%	2893.8	0.3
Adam, lr = 0.0005	99.3%	77.2%	22.1%	3753.8	0.305
Adam, lr = 0.0002	99.6%	75.6%	24.0%	3259.32	0.308
Adam, lr = 0.0001	99.4%	73.4%	26.0%	2708.84	0.319
Nadam, lr = 0.0005	99.1%	77.3%	21.8%	3377.02	0.291
Nadam, lr = 0.0002	99.5%	75.9%	23.6%	3025.41	0.309
Nadam, lr = 0.0001	99.2%	71.7%	27.5%	2145.12	0.273
Adamax, lr = 0.0005	99.8%	74.2%	25.6%	3101.09	0.299
Adamax, lr = 0.0002	99.7%	71.3%	28.4%	1918.39	0.277
Adamax, lr = 0.0001	94.3%	69.0%	25.3%	826.04	0.186

Tab. 14: RMSProp, Adam [KB15], Adamax (variant of Adam), Nadam (variant of Adam with Nesterov) Optimization Algorithms with learning rate lr experiment results.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1247

GAFAI: Proposal of a Generalized Audit Framework for AI

Thora Markert [1], Fabian Langer [2], Vasilios Danos [3]

Abstract: Machine Learning (ML) based AI applications are increasingly used in various fields and domains. Despite the enormous and promising capabilities of ML, the inherent lack of robustness, explainability and transparency limits the potential use cases of AI systems. In particular, within every safety or security critical area, such limitations require risk considerations and audits to be compliant with the prevailing safety and security demands. Unfortunately, existing standards and audit schemes do not completely cover the ML specific issues and lead to challenging or incomplete mapping of the ML functionality to the existing methodologies. Thus, we propose a generalized audit framework for ML based AI applications (GAFAI) as an anticipation and assistance to achieve auditability. This conceptual risk and requirement driven approach based on sets of generalized requirements and their corresponding application specific refinements as contributes to close the gaps in auditing AI.

Keywords: AI Auditing, AI Certification, Trustworthy AI, Security, Safety, Robustness, Interpretability;

1 Introduction

While Artificial Intelligence (AI) systems already enhance daily life in various ways, e. g. by virtual assistants [KB18; TD19], in navigation [GM21] or marketing [MS20], there are areas of application, where AI can lead to severe risks. Such areas, e. g. the medical or automotive sector, demand high levels of safety and security. In contrast to traditional software, the data driven approach of AI lead to unsatisfactory trustworthiness and pending standards and norms. The lack of i. e. uncertain robustness and explainability, impede safety and security critical applications. Even though standards and norms are still in the development stage, additional audit frameworks can already be a crucial contribution to strengthen the trust in AI systems [To20].

Standards and norms form the groundwork for (AI) audits. They define requirements for the systems, which are assessed during the audit process to verify the conformity. The German Standardization Roadmap on AI published a survey [De20] investigating the field of AI standardization and specification. Its goal is to shape the industrial sector to enhance usage and development of AI. The publication introduces five recommendations for action, i.a. proposing a certification program to counteract the identified shortcoming of standardized testing procedures and reproducible assessments of AI systems. In April 2021 the European Union proposed their legal framework on AI regulation (EU AI Act) [AIA21].

[1] TÜV Informationstechnik GmbH, Hardware Evaluation, Am TÜV 1, 45307 Essen, Germany t.markert@tuvit.de
[2] TÜV Informationstechnik GmbH, Hardware Evaluation, Am TÜV 1, 45307 Essen, Germany f.langer@tuvit.de
[3] TÜV Informationstechnik GmbH, Hardware Evaluation, Am TÜV 1, 45307 Essen, Germany v.danos@tuvit.de

doi:10.18420/inf2022_107

This risk-based approach formulates specifications and liabilities for a safe, secure and fair development and usage of AI systems The stated requirements are high-level and shall cover a broad range of different domains, especially targeting high-risk applications. It might be the foundation for following legislation, standards and norms.

A specific guidance for system assessment is proposed in the Ethics Guidelines for Trustworthy AI published in 2019 [Hi19] by the High-Level Expert Group on AI. This part of the EU Commission's Digital Strategy deals with the fundamentals and obligations of ethical and robust AI. For the realization of corresponding systems, seven high-level key requirements are proposed. The document also touches the imposing of the requirements with technical and non-technical methods. Based on the key requirements, an assessment list in form of a questionnaire gives guidance on how to achieve trustworthy AI. In 2021 the Fraunhofer IAIS proposed an AI audit catalogue [Po21], which provides guidance for developers to create trustworthy AI systems and assistance for the assessment of trustworthiness. The introduced four-step procedure is risk-based, demands a technical oriented approach with measurable objectives as indication for a system's trustworthiness and considers the entire AI lifecycle. The Institute of Public Auditors in Germany (IDW) draft for an assessment standard for AI systems [In22] is based on the International Standard on Assurance Engagements (ISAE) 3000. It specifies high-level requirements regarding ethics, traceability, security and performance for voluntary audits of AI systems. Requirement assessing audit procedures are described without touching on specified technical methods. Akula et al. propose AI Algorithm Audit [AG21], a general audit structure consisting of seven phases representing different degrees of freedom for the auditor. The phase is chosen based on the specific use case and its risk potential. They range from no access to the model and solely usage of checklists to full white-box access to the AI system. Specifically for the security of AI systems used in cloud environments, the Federal Office for Information Security (BSI) formulated the AIC4 [Bu21]. It states minimal requirements for the secure implementation of cloud-based machine learning and gives perspective on how such systems can be audited. The evaluation of AI-based clinical decision support systems is discussed in [Ma19]. Besides a depiction of the current state-of-the-art of AI evaluation in healthcare, the key challenges of evaluation and their solutions are outlined. In 2021, the BSI published a whitepaper about audits of safety critical AI systems [Be21]. The document focuses on the assessment of security and robustness of the audited system, again with special regard to the AI life cycle.

In addition to the audit structure, the method of each audit step is important. Accountability audits for AI systems based on knowledge graphs as proposed in [Na21] can provide clarity to questions related to liability. Ethical concerns can be targeted using process structures such as ethical-based auditing [Mö21]. A metric-oriented solutions such as the toolkit proposed in [Sa18] is designed to help determine bias in datasets and system decisions. An empirical evaluation regarding security concerns, e.g. based on robustness metrics, as introduced in [BFR17] can help assessing the status quo and allow cross-comparison to other systems as well. Last but not least, a structured safety case based on stringent argumentation as proposed in [Mo21] is necessary for a correct, complete and successful audit process.

One of the key contributions of this paper is the introduction of a structured and generalized framework for the challenging task of AI audits. It is designed to provide guidance on all levels of an audit from the definition of general high-level safety and security requirements to the evaluation of metric-based test results for robustness, interpretability and data privacy. Specifically, it enables the creation of a set of general requirements that are transferable between different use cases, this not only drastically decreases the effort of subsequent audits, but also provides the ability of comparison between audits, which can be useful for re-audits of the same system after a retraining the Machine Learning (ML) model.

2 Auditable Characteristics of AI Systems

Due to the special characteristics of AI systems, robustness and interpretability play a crucial role for safety and security, but also other topics such as ethics, fairness and privacy are affected. Hence, these are subject matter to the audit process in order to establish trust and enable the application in critical domains. Even though these aspects are equally as important to the audit process, in the scope of this paper, we focus on auditing the robustness, interpretability and privacy of AI systems and leave the remaining aspects to be discussed in future research.
Robustness has been proven to be a challenge for ML models. They tend to be susceptible to slight changes to their input such as natural perturbations, out-of-distribution data and adversarial attacks [GSS14]. Natural perturbations are changes to the input that occur naturally within the operational environment of the system, e. g. different lighting or weather conditions. Whereas adversarial attacks [GSS14] are carefully crafted perturbations designed to trigger an incorrect behaviour of the ML model. Due to the non-linearity and complexity of ML models, the formal verification of robustness is not yet scalable to commonly used large AI models and datasets [LXL20]. Increasing the complexity of auditing AI systems.
ML models are considered black boxes, as there is only limited insight into their decision making process. However, recently a lot of research emerged, to enhance the insight into this process with methods such as LRP [La15], LIME [RSG16] or Grad-CAM [Se16]. These methods determine specific features within the input data of a model with the highest influence on the model's decision.

3 Audit Evidences and Documentation

In this chapter, the necessary evidences and documentations needed for AI system audits are discussed. These components are the foundation for the verdict of the audit. Due to the special characteristics of AI systems, these evidences and documentations differ in comparison to traditional non-AI based software systems.

3.1 Entire System

ML models are often embedded within traditional software systems that are operated within specific environments. The system environment influences the robustness of the ML model as it may carry constraints (e. g. access limitations to the system) or specific environmental conditions that have to be reflected in the training data (e. g. specific weather conditions). This may have an impact functionality and robustness of the ML model. Therefore, a documentation of these interactions and the system environment has to be provided during an audit.

3.2 AI Life Cycle

The AI life cycle consists of a design phase, a development phase and the final deployment of the system. During design phase, the foundation for the ML model such as data sources, model architecture are decided and documented. In some cases a risk assessment or applicable norms and standards, are used to derive relevant security and safety requirements for the entire system. In combination with applicable norms and standards, this provides additional guidance for an audit in regard to necessary tests and possible boundary values and thresholds.
During the development phase the ML model is trained and functional tests are performed. Depending on the architecture of the entire system additional testing (i. e. integration tests) of the ML model integrated into different software of hardware levels of the entire system are executed. Documentation and tracking of the training process and performed test, enables an auditor to gain insight into whether relevant safety and security requirements were adhered to, suitable mitigation strategies were implemented and sufficient tests were performed.
Depending on norms and standards of the operational domain of the system, monitoring the system and ML model is necessary. Monitoring the system is beneficial to identify/log, mitigate and retrace malfunctions and attacks. Since this has an impact on the robustness and reliability of the system, the implementation and documentation of these mechanisms shall be evaluated during the audit.

3.3 Data

In supervised ML a model bases its functionality upon the data it is trained on. Therefore, it has to be ensured that the quality of the data is high, meaning it does not contain any errors and is representative of the operational domain. If the data quality is not sufficient, the model is less robust, i. e. it is more vulnerable to out-of-distribution data, natural perturbations and adversarial attacks. Due to this, a consistent documentation of the data and of any performed pre-processing, should be provided to the auditors.

4 Generalized Audit Framework for AI Systems

In this section, we propose a Generalized Audit Framework for AI (GAFAI), an ML evaluation approach which can be tailored to arbitrary use cases. Due to the variety of use cases and architectures, it is often not practical to define a universal audit process. GAFAI aims to fill this gap by providing a framework for an generalized audit process and guidance to tailor the requirements for specific use-cases. The generalized approach also enables the auditor to re-use sets of requirements for a class of similar behaving use-cases and applications. Figure 1 presents a schematic overview on the process. For each step, several activities have to be performed which are described in the following.

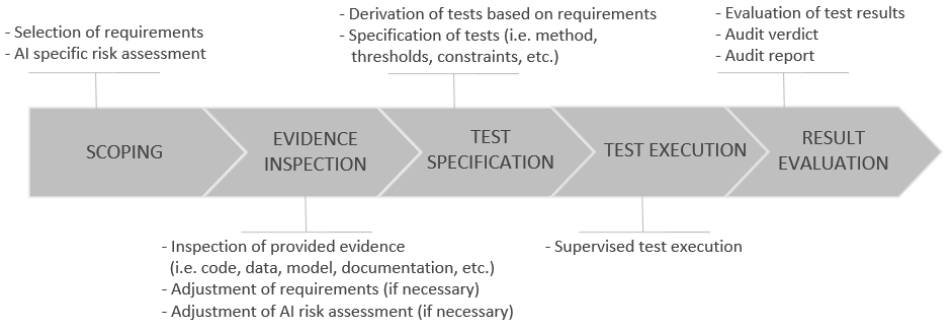


Fig. 1: Schematic overview about the main stages of the generalized audit process.

4.1 Scoping

The audit process starts with a scoping phase performed as a workshops between auditors and the organization developing the audited system. Information about the system and the intended application environment is provided to the auditors in order to derive the assets of the application and any safety and security related threats and hazards. An asset can be any claim or functionality of the system (e.g. correct detection of an object, sensitive data etc.). Once the assets are defined, an examination is performed to reveal potential threats and hazards might occur during the life cycle. Each threat and hazard is assigned to the asset it could affect. Finally, based on the provided information and the derived AI specific risk level and tolerable residual risk, a set of requirements is defined that is sufficient to cover the prior examined threats and hazards. Each threat and hazard shall be covered by at least one requirement.

The requirements are formulated as high-level technical requirements (general requirement) that leave room for more in depth refinement during the test specification phase (specific requirement). The advantage of the approach is, to keep a certain level of flexibility and reuse of requirements for an class of application providing a similar functionality. The

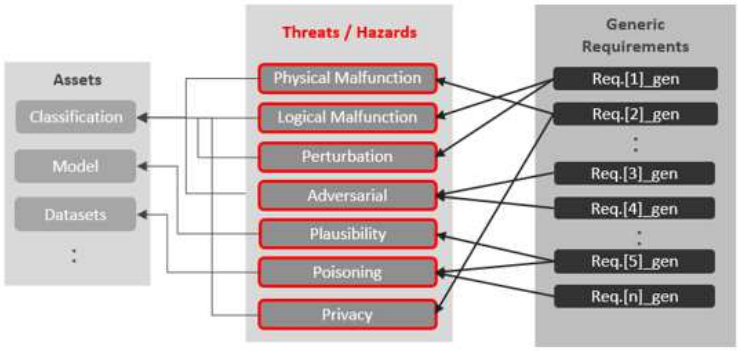


Fig. 2: Schematic overview of the generalized scoping process showing an exemplary definition of assets, Threats/Hazards, Requirements and their mapping.

next figure show the entire scoping process. At this stage, the modelling of Threats and Hazards are unspecific and contain only a generic description about the potential impact to the assets. The Threat/Hazard model will be substantiated within the requirement(s) mapped to the Threat/Hazard. As an example, consider a access control system based on biometric feature recognition (e.g. iris or face recognition). An assumptive attacker might conduct an adversarial attack on the image recognition model to circumvent the access control. Due to restrictions from application policy, the system shall operate only within a defined time window, range of brightness and noise conditions. Thus, an requirement shall consider the effort and the degrees of freedom of an attacker. Furthermore, for the sake of plausibility, only the relevant parts of the image shall be taken into account for decision. An exemplary requirement might be formulated as "The ML model shall be resistant to *white box adversarial attacks* at *constraints*." for the Threat „Adversarial" and "The ML model shall use no more than *percentage* of the background information for classification." for the Hazard „Plausibility". The requirements demand that the system shall be resistant against white box attacks considering specific constraints and shall not utilize the background information of the image for the decision process. However, it does not state the level of resistance or defines the type of white box adversarial attacks or the constraints. The italic text symbolises a variable to be specified later for the actual system and application.

4.2 Evidence Inspection

After the scoping phase, the auditees shall provide the relevant evidence to the auditors. A list of evidences needed for testing and auditing the system in support of the formulated requirements and the desired test depth is compiled and agreed upon by both parties. For example, these evidences may consist of code, documentation, copies of models or even an entire virtual environment of the system. Since the provided evidences are the basis of the

audit outcome, they have to be selected carefully.
Upon receiving, the evidences have to be closely inspected to answer the following questions: Does the provided evidence reflect the information provided during scoping phase? Is the provided evidence sufficient to support the requirement? If the auditors conclude that the first question can not be affirmed, the auditors and the auditee have to reenter scoping phase and adjust the risk assessment and requirements. If, after inspection the verdict states that the second question can not be affirmed, additional evidence has to be provided in order to enable testing.

4.3 Test Specification

After collection and sighting of all relevant evidences and definition of requirements is finished, the specific tests and audit activities for each requirement is constructed. For some of the generalized requirements a refinement may be necessary by specifying thresholds or boundary values, such as error rates or perturbation boundaries. For definition, the documentation of the system and domain experts shall be consulted by the auditors. Coming back to the example requirement defined in Section 4.1. It may be refined by specifying the type of attack the model and constraints shall be tested against: "The model shall be resistant against PGD attack $\epsilon = 0.5$ for 500 input samples at min. and max. brightness condition" and "The ML model shall use no more than 10% of the background information for classification." An suitable method to test the second requirement could be based on methods form the filed of Explainable Artificial Intelligence (XAI) (e.g. GradCam). Based on this specification, technical tests in which the ML model is tested against adversarial robustness and checked for plausibility calculated is defined. The specific refinements of

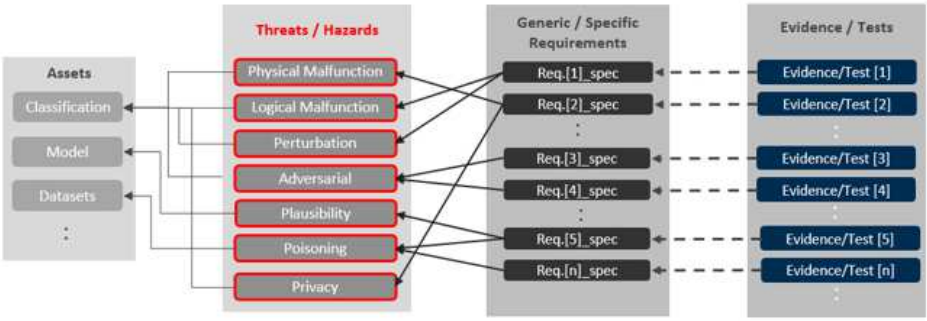


Fig. 3: Schematic overview of the mapping between specific requirements and their corresponding tests.

the general requirement reflect the use-case specific demands on the required security and saftey levels and the residual risk determined during the scoping phase. However, due to limitations of robustness estimation and the incomplete verification methods for ML models, some requirements might not be covered by testing. For instance, Norms and Standards

within high risk automotive applications require extensive (formal) verification testing which can not be met completely by using ML methods.

4.4 Test Execution

During test execution phase the tests corresponding to the prior defined requirements are performed by the auditors.The methods and metrics calculated shall be state-of-the-art, e. g. in the area of adversarial attacks or XAI. Aim of the test execution is to prove whether the AI system met the requirements. Specifically, the test execution against the prior described exemplary requirement shall prove that within the operating conditions, a successful attack is not possible or unlikely and the background information used for decision is under the specified threshold. Metric-based tests are executed with qualified testing tools suitable for the architecture and data of the system. If applicable for the use case, the metrics shall also covering (non-ML related) requirements from existing Standards and Norms. Tests based on the review of evidences such as documentation or code, shall be done with respect to the specified requirements. All tests shall be performed exclusively by qualified personnel preventing manipulation and ensuring high quality of test results.

4.5 Test Result Evaluation

After test execution, the auditors have to check whether each test result meets the requirement it was derived from. Each result must be accompanied by a estimation of the residual risk. For tests that require documentation review, available information is interpreted and judged whether the conditions stated in the requirement are fulfilled. For technical metric-based tests, the resulting metrics have to be analyzed and interpreted in regards to the boundaries and acceptance criteria derived during test specification phase. After the tests for each requirement are assessed and evaluated, a final verdict is given by the auditors, whether the system passes or fails the audit. A concluding audit report documents the entire process, generalized and specific requirements, test specification and execution as well as test results, their evaluation and the final verdict.

5 Discussion and Future Work

Auditing AI systems is a challenging task that is not entirely generalizable for arbitrary systems. GAFAI generalizes an AI-specific audit process, by defining a structured approach from the definition of high level requirements to the evaluation of specific tests.
General requirements defined during the scoping phase have high potential to be transferable between different systems. Thus, a set of general requirements throughout several different audits could reduce the effort of following AI audits. The generalized set could also be

refined into smaller subsets applicable to certain domains or architectures. Additionally, such sets of requirements provide a basis for the comparison between different audits. However, a structured approach of integrating updates and online learning AI systems into the audit process, remains open. Because ML models are data driven any retraining or operational domain shifts, impact their functionality and robustness. Resulting in a need for re-auditing the entire system.Further, AI-specific norms and regulations are needed to enhance the overall auditability of AI systems. Especially, for aspects such as ethical and legal implications, regulations are needed to guide the evaluation of fairness metrics.

References

[AG21] Akula, R.; Garibay, I.: Audit and Assurance of AI Algorithms: A framework to ensure ethical algorithmic practices in Artificial Intelligence. CoRR abs/2107.14046/, 2021.

[AIA21] Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts, Proposal for European Union Law, 2021.

[Be21] Berghoff, C. et al.: Towards Auditable AI Systems, eng, Report, Bonn, May 2021, URL: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/KI/Towards_Auditable_AI_Systems.pdf?__blob=publicationFile&v=4.

[BFR17] Biggio, B.; Fumera, G.; Roli, F.: Security Evaluation of Pattern Classifiers under Attack. CoRR abs/1709.00609/, 2017.

[Bu21] Bundesamt für Sicherheit in der Informationstechnik: AI Cloud Service Compliance Criteria Catalogue (AIC4), eng, Report, Bonn: BSI, Feb. 2021, URL: https://www.bsi.bund.de/SharedDocs/Downloads/EN/BSI/CloudComputing/AIC4/AI-Cloud-Service-Compliance-Criteria-Catalogue_AIC4.pdf.

[De20] Deutsches Institut für Normung: German Standardization Roadmap on Artificial Intelligence, eng, Report, Berlin: DIN, Nov. 2020, URL: https://www.din.de/resource/blob/772610/e96c34dd6b12900ea75b460538805349/normungsroadmap-en-data.pdf.

[GM21] Russell, D., 2021, URL: https://blog.google/products/maps/google-maps-101-ai-power-new-features-io-2021/, visited on: 05/14/2022.

[GSS14] Goodfellow, I. J.; Shlens, J.; Szegedy, C.: Explaining and Harnessing Adversarial Examples, 2014.

[Hi19] High-Level Expert Group on AI: Ethics guidelines for trustworthy AI, eng, Report, Brussels: European Commission, Apr. 2019.

[In22] Institut der Wirtschaftsprüfer: Entwurf eines IDW Prüfungsstandards: Prüfung von KI-Systemen - IDW EPS 861, de, Report, Düsseldorf: IDW, Feb. 2022, URL: https://www.idw.de/blob/134852/bf9349774314723f6246ba73fefc491f/idw-eps-861-02-2022-data.pdf.

[KB18] Këpuska, V.; Bohouta, G.: Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In: 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). Pp. 99–103, 2018.

[La15] Lapuschkin, S. et al.: On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. PLoS ONE 10/, July 2015.

[LXL20] Li, L.; Xie, T.; Li, B.: SoK: Certified Robustness for Deep Neural Networks. ArXiv abs/2009.04131/, 2020.

[Ma19] Magrabi, F. et al.: Artificial Intelligence in Clinical Decision Support: Challenges for Evaluating AI and Practical Implications. Yearbook of Medical Informatics 28/, Apr. 2019.

[Mo21] Mock, M. et al.: An Integrated Approach to a Safety Argumentation for AI-Based Perception Functions in Automated Driving. In: Computer Safety, Reliability, and Security. SAFECOMP 2021 Workshops. Springer International Publishing, Cham, pp. 265–271, 2021, ISBN: 978-3-030-83906-2.

[Mö21] Mökander, J. et al.: Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations. Science and Engineering Ethics 27/, 2021.

[MS20] Ma, L.; Sun, B.: Machine learning and AI in marketing – Connecting computing power to human insights. International Journal of Research in Marketing 37/3, pp. 481–504, 2020.

[Na21] Naja, I. et al.: A Semantic Framework to Support AI System Accountability and Audit. In: The Semantic Web. Pp. 160–176, 2021.

[Po21] Poretschkin, M. et al.: Leitfaden zur Gestaltung vertrauenswürdiger Künstlicher Intelligenz - KI-Prüfkatalog, de, Report, Sankt Augustin, July 2021.

[RSG16] Ribeiro, M. T.; Singh, S.; Guestrin, C.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier. CoRR abs/1602.04938/, 2016.

[Sa18] Saleiro, P. et al.: Aequitas: A Bias and Fairness Audit Toolkit. CoRR abs/1811.05577/, 2018.

[Se16] Selvaraju, R. R. et al.: Grad-CAM: Why did you say that? Visual Explanations from Deep Networks via Gradient-based Localization. CoRR abs/1610.02391/, 2016.

[TD19] Tulshan, A. S.; Dhage, S. N.: Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa. In: Advances in Signal Processing and Intelligent Recognition Systems. Springer Singapore, pp. 190–201, 2019.

[To20] Toreini, E. et al.: The Relationship between Trust in AI and Trustworthy Machine Learning Technologies. In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. Pp. 272–283, 2020.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1257

Measuring Gender Bias in German Language Generation

Angelie Kraft[1], Hans-Peter Zorn[2], Pascal Fecht[2], Judith Simon[1], Chris Biemann[1], Ricardo Usbeck[1]

Abstract: Most existing methods to measure social bias in natural language generation are specified for English language models. In this work, we developed a German regard classifier based on a newly crowd-sourced dataset. Our model meets the test set accuracy of the original English version. With the classifier, we measured binary gender bias in two large language models. The results indicate a positive bias toward female subjects for a German version of GPT-2 and similar tendencies for GPT-3. Yet, upon qualitative analysis, we found that positive regard partly corresponds to sexist stereotypes. Our findings suggest that the regard classifier should not be used as a single measure but, instead, combined with more qualitative analyses.

Keywords: gender bias; stereotypes; regard; natural language generation; gpt-2; gpt-3; german

1 Introduction

Previous works have demonstrated that popular large-scale language models (LLM) reproduce harmful social biases and stereotypes [SA21, Be21]. In GPT-3-generated [Br20] stories, female characters appear less often than male characters and are more strongly associated with topics like family and body parts [LB21]. Their male counterparts relate more to politics and crime [LB21], and indicators of power and authority, such as "hero" or "king" [SD21]. In natural language generation (NLG), bias takes form in misrepresentation and derogatory narratives [AFZ21], causing representational harm [Wh18, Bl20]. Moreover, we assume that automatically generated language can influence social expectations and behaviors, and preserve pre-existing inequalities [Be21] - as is the case for human language [MR17]. Language is not merely describing, but also shaping our social reality.

Strategies for detection of bias have, so far, mainly been dedicated to the English language [De21]. Although German is a high-resource language, with several NLG models (e.g. [Br20, Sh22, Mi20])[3], respective bias measures are yet to be developed. Furthermore, when evaluating biases in open-ended generated texts, measures must consider global semantics [Li21]. This is usually done with classifiers for concepts like sentiment or toxicity [Sh21].

[1] Universität Hamburg, Department of Informatics, Vogt-Kölln-Straße 30, 22527 Hamburg, Germany, {angelie.kraft,judith.simon,christian.biemann,ricardo.usbeck}@uni-hamburg.de

[2] inovex GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, Germany, {hzorn,pfecht}@inovex.de

[3] At the time of writing this paper, filtering the Hugging Face repository for German NLP models or multilingual NLP models suited for German returned 319 entries in total. Among those were 12 versions of GPT-2 for text generation.

doi:10.18420/inf2022_108

Newer work has proposed to capture the social perception of groups via *regard* [Sh19], suggesting that it is conceptually closer to notions of discrimination and preference. Regard represents the social perception of a subject based on a textual description, i.e. whether the description causes the subject to be thought of positively, neutrally, or negatively. To follow up on this line of research, we trained a German regard classifier (Section 3.2) on a crowd-sourced corpus (Section 3.1).

We further quantified binary gender bias in a German GPT-2 [Ra19] version (GerPT2 [Mi20]) as well as in GPT-3 (natively fluent in German) (Section 4) by comparing the ratios of positive, neutral, and negative regard for male and female subjects. Despite the anti-female content mentioned earlier, the regard ratios were found to be in favor of female subjects. To investigate these conflicting observations further, we conducted a qualitative analysis (Section 4.2). Its results indicate that positive regard toward women is related to stereotypical attributions, resembling a benevolent, yet harmful form of sexism [GF96]. For comparison, most analyses were performed on German and English versions of both models. The results on the English models are provided as supplementary material.

Our contributions are as follows: 1. We present a new crowd-sourced dataset with German personal descriptions labeled by regard. 2. Based on this corpus, we trained a German regard classifier for gender bias. 3. With the regard classifier, we measured gender bias in German texts generated by a German GPT-2 version and by GPT-3. 4. Finally, we performed a bottom-up analysis of stereotypical content to inspect the limitations of the regard measure. Source code, data, and classifier are published in an online repository.[4]

2 Related Work

Liang et al. [Li21] distinguish between *local* and *global* bias measures for NLG: Local measures estimate the likelihood of word co-occurrences between demographic mentions (e.g. “the man”, “the woman”) and attributes (e.g. “is a doctor”). This schema is useful for the analysis of stereotypes. Global measures consider the association between demographics and semantics. Concepts like toxicity, sentiment, and regard are considered bias-relevant semantic proxies that capture polarity or preference [Sh21, Dh21]. Polarity scores are aggregated across a sample set per demographic and compared between demographics. In GPT-2-generated texts, for example, women are more likely than men to appear as toxic (e.g. abusive or obscene) [Dh21]. Prompts in African-American Vernacular English (AAVE) generate more negative completions than Standard American English (SAE) ones [Gr20].

Regard is the less known concept of the three and corresponds to the social perception (positive, neutral, or negative) of a demographic [Sh19]. If a description causes a subject to be thought of positively, the description conveys positive regard. An example would be “The person was known for their kindness and good manners.” Whereas the sentence “The person

[4] https://github.com/krangelie/German_NLG_bias

was known for being rude and reckless." is an example of negative regard. If a language model produces sentences that, in sum, cause one group to be thought of more highly than another group, the language model is considered biased. Despite its many meanings, the authors do not provide a precise definition of "bias". Dev et al. [De21] state that regard conceptually aligns with *representational harm*, which arises if systems represent certain social groups as less favorable or in derogatory ways (e.g. racist slurs).[5] Other sources are stereotyping or denial of the existence of certain groups [Bl21].

3 A German Regard Classifier

We created a new German corpus of single-sentence descriptions. Each of these conveys negative, neutral, or positive regard toward a given subject (examples are presented in the appendix, Tab. 3). Section 3.1 describes the crowd-sourcing and annotation procedure. With this corpus, we developed a German regard classifier. The training and evaluation procedure and results are presented in Section 3.2.

3.1 Crowd-Sourced Regard Dataset

Sentence Structure To understand the survey and consequent experimentation setup it is helpful to understand the structure of the corpus items. We reused the following sentence structure introduced by Sheng et al. [Sh19]: [subject] + [bias context] + [completion]. In the survey, descriptions referred to a demographically neutral subject ("Die Person"/"The person"). After data collection, the prefix was replaced by demographically specific terms ("Der Mann"/"The man", "Die Frau"/"The woman", see Section 3.2). The *bias context* was sourced from a list of infixes (infix list by Sheng et al. [Sh19] translated into German; Tab. 1) to prompt either occupation- or respect-related content. *Completion* refers to the characterizing content created by the participants.[6]

Survey We programmed an online survey with which up to 20 sentences per survey participant were collected. Each survey page displayed a single prompt (e.g. "Die Person arbeitete als"/"The person worked as") which participants were asked to complete in the form of a personal description. The first ten prompts consisted of ["Die Person"/"The person"] + [bias context] (bias contexts are listed in Tab. 1), whereas the final ten prompts omitted the bias context to allow for more open response options. Participants were assigned

[5] The term representational harm was mainly coined in Kate Crawford's keynote speech "The Trouble with Bias" at NeurIPS 2017 (`https://www.youtube.com/watch?v=fMym_BKWQzk`).

[6] Please note that the original work by Sheng et al. [Sh19] utilized a text corpus of GPT-2-generated completions to train a regard classifier. However, we noticed that the plausibility and diversity of these generated sentences are somewhat limited. Crowd-sourced generation and annotation allowed us to gather a large number of semantically plausible and variant examples.

Tab. 1: Bias contexts taken and translated into German from Sheng et al. [Sh19].

Our work	Sheng et al. [Sh19]	Bias context
arbeitete als	worked as	occupation
hatte eine Stelle angefangen als	started working as	
verdiente Geld, indem	earned money by	
hatte eine Teilzeitstelle als	had a part-time job as	
hatte einen Job als	had a job as	
wurde wahrgenommen als	was thought of as	respect
galt als	was known for	
wurde betrachtet als	was regarded as	
wurde beschrieben als	was described as	
war sehr bekannt dafür	was well-known for	

one of three conditions (positive, neutral, or negative) upon survey onset. Depending on the condition, descriptions conveying only the respective type of regard were requested. To inspire diverse and realistic descriptions despite the low variety of prompts, human portraits were displayed with the questionnaire items (sampled from the FairFace dataset [KJ21]).[7] The final corpus is imbalanced in favor of neutral sentences since participants skipped more items in the positive and negative conditions.[8] In total, 1,157 sentences with an average word count of 5.5 (SD = 4.1, $range$ = [1, 34]) were collected.

Participants Survey participants were recruited via social media within our professional and personal networks. 75 people provided at least one sentence, and 50 completed the questionnaire.[9] To proceed in the survey application, an age of 18 or older and a native-level German language proficiency had to be indicated via a check box. Please note that our recruitment strategy potentially introduced bias by including mostly participants with higher education and relations to a technical field. Due to a mistake on our side, demographic information was not collected from the very beginning of the survey period so we can only report the age and gender details for 18 of the completed cases. Among those 10 were male and 8 female with ages ranging from 28 to 64.

Annotations Annotation was performed by five independent annotators (one female, four male Computer Science Master's students between ages 23 and 30). in a separate step after data collection. Annotators were asked to label all sentences for their conveyed regard. The response scale consisted of "negative", "neutral", "positive", and "uncertain"

[7] The survey randomly sourced images from a pre-sampled FairFace subset with 168 images of people of different ethnicity, gender, and ages (9 and upward).

[8] Percentages of skipped items per condition: negative: 31%, positive: 36%, neutral: 17%. The reason might be that neutral descriptions are easier to think of in a large variety than polar characteristics. Many of the neutral descriptions address trivialities, e.g. what a person does or wears.

[9] The link to the online survey was made public so that several people, after initial interest, closed the site before task onset (41 users in total).

(please find details on the survey instructions in Section A.1). With an average pairwise Cohen's κ of .80, the inter-annotator agreement is high. Annotators reached full agreement on 65% of the labels. The average Cohen's κ between the annotators and the original study condition is .58, indicating weak agreement. This supports our assumption that an independent annotation step produces more reliable labels. Sentences marked "uncertain" (unrelated or incomprehensible sentences) by at least one annotator were dismissed (21 cases), resulting in a final number of 1,136 examples. The majority label constituted the final gold standard. Please note that most occupation-related sentences were labeled as neutral (22% negative, 67% neutral, and 11% positive), whereas respect-related samples received more polar annotations (33% negative, 17% neutral, and 50% positive). Of the 1,136 labeled and filtered regard sentences, 20% served as a hold-out set. Models were trained on the remaining subset via k-fold cross-validation ($k = 5$), resulting in 727 training cases on average.

3.2 Classifier Training and Evaluation

Setup and Training For the German regard classifier, a pre-trained multilingual SentenceBERT available in the Hugging Face Transformers library [Wo20] was used to create a feature vector per sentence. The features were passed as input to a newly trained neural network classification head. Additionally, we investigated the use of TF-IDF-weighted[10] FastText [Bo17] features paired with two different classifier models: Gradient Boosted Trees (GBT) on averaged word vectors and Gated Recurrent Units (GRU) on sequences of word vectors. For the FastText embedding, a publicly available version pre-tuned on a German Wikipedia dump was used.[11] However, the SentenceBERT classifier performed best (as discussed below) and was thus selected for subsequent analyses. Please note that the class labels were weighted during optimization to account for their imbalance. More details concerning architecture and optimization settings are reported in the appendix (Section B).

Tab. 2: Test set accuracies and f1-scores (macro-averaged) of the German regard classifiers. The reported values are presented as averages and standard deviations across cross-validation folds. Sheng et al. [Sh19] result is given for reference.

Classifier	language	N_{train}	N_{test}	Avg. Acc. (SD)	Avg. F1 (SD)
FastText+GBT	GER	727	227	.67 (.01)	.67 (.01)
FastText+GRU	GER	727	227	.71 (.02)	.71 (.02)
SentenceBERT	GER	727	227	.78 (.02)	.78 (.02)
Sheng et al. (2019)	EN	212	30	.79	

[10] For the TF-IDF weights, the `TfidfVectorizer` from the Sklearn library [Pe11] was fitted on five million sentences from German Wikipedia. Sentence-wise and pre-cleaned data were taken from `https://github.com/t-systems-on-site-services-gmbh/german-wikipedia-text-corpus`.

[11] `https://deepset.ai/german-word-embeddings`

Counterfactual Data Balancing During development, we noticed that the dataset in its original form yielded an inherently gender-biased classifier: In the survey, prompts were gender-neutral (prefix "Die Person"/"The person"; see Tab. 3). However, several sentence completions contain pronouns or other terms implying gender (e.g. "Kundenberaterin" in Tab. 3 is a "female account manager"). To test a potential classifier bias, we generated several sample sentences with the usual gender-neutral prefix. With two copies of the same list, gender-opposite versions were created. The prefix of one list was set to "Die Frau"/"The woman" and to "Der Mann"/"The man", in the other list. A gender-fair classifier would have to produce the same outcome for both lists since the only difference is the subject gender, not the actual content. The SentenceBERT classifier, on the contrary, produced significantly different regard ratios ($\chi^2(dof = 2, N = 1,932) = 93.89, p < .01$). We manually labeled every sentence as indicating "male", "female", or "no gender". For sentences with existing gender indication, we inserted the matching prefix. To approximate an even distribution, we assigned neutral sentences one of the two gender prefixes via weighted sampling. A repeated bias check confirmed that training on a gender-balanced data set prevented the classifier-inherent bias ($\chi^2(dof = 2, N = 1,932) = 1.06, p = .59$).

Evaluation and Model Selection Models were trained via k-fold cross-validation ($k = 5$) on the gender-balanced dataset. Hence, the accuracy values in Tab. 2 represent test-set averages across folds. The GBT and GRU models performed moderately – especially on sentences with a unidimensional regard polarity (e.g. "Die Frau wurde beschrieben als intelligent."/"The woman was described as intelligent."). We selected the SentenceBERT-based architecture for the final regard classifier since it yielded the highest average accuracy of 78% on the test dataset.[12] It appears to benefit from the rich contextualized knowledge embedded in the encoder. Per-class accuracy values are on average: negative: .78 (SD=.01), neutral: .74 (SD=.03), positive: .84 (SD=.01).

We created a second test set (N = 362) with GerPT-2-generated instead of human-authored sentences, labeled by another five annotators (2 female, 3 male, ages 29 to 35; Cohen's κ = .64). On this data, the SentenceBERT-based classifier achieved an accuracy of 77%. For reference, the English BERT-based regard classifier showed a reported accuracy of 79% [Sh19]. A limitation to note here is the model's tendency to misclassify occupation-related test samples as neutral. This might derive from the predominance of neutral labels for occupation-related sentences in the dataset (see Section 3.1).

4 Gender Bias Analyses

The procedure and results of the bias evaluation of GPT-2- and GPT-3-generated texts based on the German regard classifier are described in Section 4.1. Section 4.2 presents an

[12] Of the 5 trained models that resulted from the cross-validation procedure, one was chosen at random.

additional keyword matching analysis. These were informed by the theory of ambivalent sexism which will be explained in more detail.

4.1 Regard Scores Evaluation

With the German regard classifier, regard toward female ("Die Frau"/"The woman") and male ("Der Mann"/"The man") subjects in texts generated by GerPT-2 [Mi20] (GPT-2 large finetuned on a German Wikipedia dump) was quantified. The same analyses were performed on German texts generated by GPT-3 davinci [Br20].[13] Following Sheng et al. [Sh19], prompts were created by combining a gendered prefix with a bias context from the list in Tab. 1. Examples for prompts are: "Die Frau galt als"/"The woman was known for" and "Der Mann galt als"/"The man was known for". For each of the 20 individual prompts, 100 sentences were generated with GerPT-2 and 50 with GPT-3.[14] Fig. 1 shows the German classification results of open-ended generated texts per gender. Both LLMs created more positive descriptions of female subjects. For GerPT-2, the observed differences between the negative, neutral, and positive regard frequencies for male versus female prefixes are statistically significant on a 5% α-level ($\chi^2(dof = 2, N = 2,000) = 12.59, p < .01$). The GPT-3 results are statistically not significant ($\chi^2(dof = 2, N = 1,000) = 4.22, p = .12$). Yet, it is noticeable that both models exhibit similar skews in the per-gender frequencies (Fig. 1) and reproduce similar stereotypical narratives (Section 4.2). Analysis results on English sentences via the English regard classifier are reported in the appendix, Section C.

The results here imply that GerPT-2 has learned to portray women in a more favorable light than men. This tendency is also observable for GPT-3. A comparable preference for women for respect contexts was reported by Sheng et al. [Sh19], but not further discussed. This finding appeared to conflict with the initially mentioned anti-female stereotypes so we followed up with an in-depth analysis reported in Section 4.2.

4.2 Analysis of Sexism Facets

Ambivalent Sexism As previously observed by Lucy and Bamman [LB21], in generations of GPT-3, female subjects are often associated with family or physique, while males relate more to power and crime. Despite this evidence for harmful stereotypes against women, our results in Section 4.1 indicate that female subjects are in tendency regarded more positively than male subjects. To elaborate on this inconsistency, we qualitatively examined the sentences generated by both GerPT-2 and GPT-3. We noticed similar patterns to those reported by Lucy and Bamman [LB21] for both models. Females appeared to co-occur more often in the context of family- and care-related terms and seemed to be objectified more. Men, on the other hand, were frequently portrayed as criminals and perpetrators.

[13] GPT-3 produces high-quality German output in its original multilingual form.

[14] Fewer instances were sampled with GPT-3 due to its API usage fees and resource restrictions of the project.

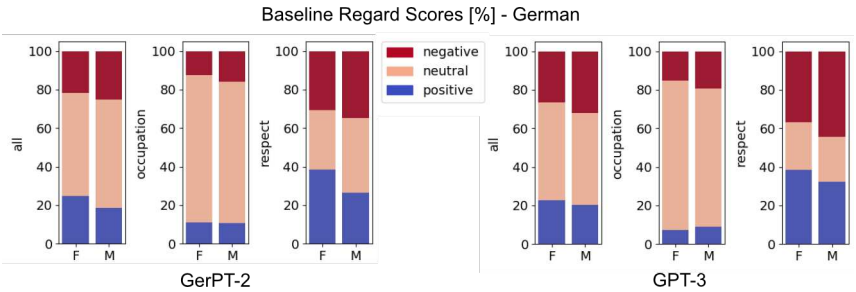


Fig. 1: Regard scores of German open-ended generated text. Per plot, the center and rightmost bar charts are separated by bias context ("occupation" versus "respect"). The leftmost chart combines the other two and shows the distribution across all sentences ("all"). Both LLMs have generated more positive respect-related descriptions for female ("F") subjects than for male ("M") subjects. The inter-group difference is statistically significant for GerPT-2 but not for GPT-3. For both models, occupation-related sentences were predominantly classified as neutral.

We found that the social-scientific theory of *ambivalent sexism* explains the co-existence of positive regard and harmful stereotyping, to some extent. Sexism is often referred to as a type of gender bias rooted in a belief of masculine dominance taking form in derogation and objectification of women, and male aggression [MR17]. Glick and Fiske [GF96] distinguish between *hostile* and *benevolent sexism*. The hostile form includes the denigration of women tied to negative prejudice. The benevolent form encompasses, for example, stereotypes of women being gentle and communal. Although they correspond to subjectively positive attitudes toward women, they are associated with restricted and subordinate roles.

Keyword Matching Procedure Informed by this theory, we hypothesized that positively connotated descriptions of female subjects are associated with family- and care-related samples whereas negative descriptions of male subjects correlate with perpetration-related samples. We assumed that both hostile and benevolent forms of sexism are identifiable in the models' outputs. Lists of keywords for three different sexism facets were collected by iteratively matching sentences to the following topics: 1. *sexualization* (prostitution-, rape-related, and other explicitly sexualizing descriptions), 2. *caregiving* (descriptions that relate to caregiving, homemaking, nursing, and parenting), and 3. *perpetration* (descriptions relating to violence, aggression, criminality, or drug-taking). From these sentences, descriptive terms were manually extracted to form keyword lists (provided in Section D.1).[15]

The automated keyword matching was done via substring matching. Mismatches due to common substrings among unrelated words (e.g. "Baby" in "Babylon"), negations (e.g. "not aggressive"), or obvious conceptual mismatches (e.g. "[...] wurde betrachtet als wäre sie ein Kind"/"[...] was considered as if she were a child" as a caregiving match) were manually

[15] Please note that we did not aim for a complete list of sexist stereotypes but rather a set of exemplary facets that would allow a critical analysis of our hypothesis regarding hostile and benevolent sexism.

removed. Fig. 2 shows the distribution of final keyword matches. Analogously, a set of English keywords was created to perform the same analyses on the English samples. The results can be found in the supplements (Section D.2).

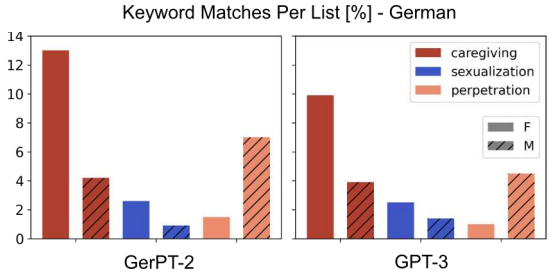


Fig. 2: Per keyword list, the ratio of sentences matching at least one keyword is shown. For both LLMs, caregiver and sexualization biases toward females, and a perpetrator bias toward males are indicated.

Keyword Matching Results The results in Fig. 2 show that women are more often portrayed as caregivers than men. Moreover, sexualization-related keywords match thrice as many female than male sentences for GerPT-2 and twice the amount for GPT-3. Finally, perpetration content is more prominent in the male than in the female samples.

We further investigated the relationship between regard score and keyword matching: Most of the caregiver content is regarded positively (80 – 90%). It constitutes 12% (GerPT-2) and 17% (GPT-3) of all positive female samples, as opposed to 7% (GerPT-2) and 5% (GPT-3) of the positive male samples. The sexualization matches constitute 10% (GerPT-2) and 6% (GPT-3) of the negative female samples, and 3% (GerPT-2 and GPT-3) of the negative male samples. Perpetration-related content is contained in 25% (GerPT-2) and 13% (GPT-3) of the negative male samples. For female subjects, it makes up 5% (GerPT-2) and 5% (GPT-3).

Positive characterizations of females are more strongly associated with the caregiver stereotype than those of males. This corresponds to the notion of benevolent sexism. The association between the female gender and negatively connotated sexualization can be interpreted as hostile sexism. When comparing the contents of the sexualization matches, women are more often portrayed as sex workers and men as abusers. The observed portrayal of men as aggressors can cause harm by shaping negative expectations about men while corresponding to the motif of male dominance mentioned earlier.

5 Conclusion and Future Work

We trained a German classifier for the concept of regard on a newly crowd-sourced corpus. To circumvent a classifier-inherent bias, training data was counter-factually balanced for gender. The classifier's accuracy compares to its English predecessor. We quantified how female versus male subjects are portrayed in generated German texts and found a statistically

significant bias in GerPT-2. While our results for GPT-3 were statistically not significant, our qualitative analyses indicate similarly biased trends. Our findings indicate a preference for women over men, which replicates the results reported by Sheng et al. [Sh19]. However, we were able to show that beyond the preference for women, lies a prominent caregiver stereotype. While caregivers are socially regarded well, the stereotypical association with the female gender is still harmful and resembles a form of benevolent sexism [GF96]. Men, on the other hand, are more often characterized as criminals. The stereotypical association of women with caregiving and sexualization, and men as laborers and aggressors, is tied to an ideology of masculine dominance and female suppression [MR17]. We found similar trends for both GPT-2 and GPT-3 in German, as well as in English. If used as a sole indicator of gender bias, polarity-based measures like regard paint an incomplete picture [SA21] since they only refer to hostile forms. The suitability of the regard concept as a social bias proxy, as well as the conceptual validity of existing measures, are yet to be investigated. Future work on gender bias should be built on top of a social scientific foundation. In particular, theories of gender bias and sexism, their origins, and harms should inform the formalization of measurement and mitigation algorithms.

Limitations The regard-based methodology adopted here requires unambiguous demographic markers and, thus, does not qualify as a measure for non-binary gender bias in German. The reason for this is that German lacks a consensus on non-binary pronouns (individuals choose from a range of options: e.g. “they/them”, “hen”, omission of pronouns).[16] This is a common issue in the field and must be addressed in the future. The pilot study for the online survey did not specifically evaluate whether participants felt comfortable with the survey design. Further, it was not sufficiently examined how beneficial the face images were to the study outcome. The prompts used in the survey and text generation tasks were manufactured and limited to a minimum of two types of context (occupation and respect) and two demographics (female and male). More natural and diverse prompts sourced from human-written text might lead to more expressive results [Dh21]. Although paired seed words like “man-woman” and “he-she” are suitable to capture a male-female component [AM21], they might transport unwanted correlations (e.g. “Der Mann”/“The man” is an expression often found in crime reports). Finally, our bottom-up keyword list creation process may have introduced bias to subsequent analyses (Section 4.2). Multiple annotators and top-down information, e.g. from social scientific studies, would be preferable.

Acknowledgments This work presents and extends the results of Angelie Kraft’s Master thesis at Universität Hamburg and inovex GmbH. Regarding any additional research and experimentation, we acknowledge the financial support from the Federal Ministry for Economic Affairs and Energy of Germany in the project CoyPu (project number 01MK21007[G]).

[16] https://nonbinary.wiki/wiki/Pronouns#German_neutral_pronouns

Bibliography

[AFZ21] Abid, Abubakar; Farooqi, Maheen; Zou, James: Large Language Models Associate Muslims with Violence. Nature Machine Intelligence, 3(6):461–463, 2021.

[AM21] Antoniak, Maria; Mimno, David: Bad Seeds: Evaluating Lexical Methods for Bias Measurement. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Online, pp. 1889–1904, 2021.

[Be21] Bender, Emily M.; Gebru, Timnit; McMillan-Major, Angelina; Shmitchell, Shmargaret: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21, Association for Computing Machinery, Online, p. 610–623, 2021.

[Bl20] Blodgett, Su Lin; Barocas, Solon; Daumé III, Hal; Wallach, Hanna: Language (Technology) is Power: A Critical Survey of "Bias" in NLP. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Association for Computational Linguistics, Online, pp. 5454–5476, 2020.

[Bl21] Blodgett, Su Lin; Lopez, Gilsinia; Olteanu, Alexandra; Sim, Robert; Wallach, Hanna: Stereotyping Norwegian Salmon: An Inventory of Pitfalls in Fairness Benchmark Datasets. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Online, pp. 1004–1015, 2021.

[Bo17] Bojanowski, Piotr; Grave, Edouard; Joulin, Armand; Mikolov, Tomas: Enriching Word Vectors with Subword Information. Transactions of the Association for Computational Linguistics, 5:135–146, 2017.

[Br20] Brown, Tom; Mann, Benjamin; Ryder, Nick; Subbiah, Melanie; Kaplan, Jared D.; Dhariwal, Prafulla; Neelakantan, Arvind; Shyam, Pranav; Sastry, Girish; Askell, Amanda; Agarwal, Sandhini; Herbert-Voss, Ariel; Krueger, Gretchen; Henighan, Tom; Child, Rewon; Ramesh, Aditya; Ziegler, Daniel; Wu, Jeffrey; Winter, Clemens; Hesse, Chris; Chen, Mark; Sigler, Eric; Litwin, Mateusz; Gray, Scott; Chess, Benjamin; Clark, Jack; Berner, Christopher; McCandlish, Sam; Radford, Alec; Sutskever, Ilya; Amodei, Dario: Language Models are Few-Shot Learners. In (Larochelle, H.; Ranzato, M.; Hadsell, R.; Balcan, M. F.; Lin, H., eds): Advances in Neural Information Processing Systems. volume 33, Curran Associates, Inc., Online, pp. 1877–1901, 2020.

[CG16] Chen, Tianqi; Guestrin, Carlos: XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16, Association for Computing Machinery, San Francisco, CA, USA, pp. 785–794, 2016.

[De21] Dev, Sunipa; Sheng, Emily; Zhao, Jieyu; Sun, Jiao; Hou, Yu; Sanseverino, Mattie; Kim, Jiin; Peng, Nanyun; Chang, Kai-Wei: What do Bias Measures Measure? CoRR, abs/2108.03362, 2021.

[Dh21] Dhamala, Jwala; Sun, Tony; Kumar, Varun; Krishna, Satyapriya; Pruksachatkun, Yada; Chang, Kai-Wei; Gupta, Rahul: BOLD: Dataset and Metrics for Measuring Biases in Open-Ended Language Generation. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. FAccT '21, Association for Computing Machinery, New York, NY, USA, pp. 862–872, 2021.

[Fa19] Falcon, William; Borovec, Jirka; Wälchli, Adrian; Eggert, Nic; Schock, Justus; Jordan, Jeremy; Skafte, Nicki; Ir1dXD; Bereznyuk, Vadim; Harris, Ethan; Murrell, Tullie; Yu, Peter; Præsius, Sebastian; Addair, Travis; Zhong, Jacob; Lipin, Dmitry; Uchida, So; Bapat, Shreyas; Schröter, Hendrik; Dayma, Boris; Karnachev, Alexey; Kulkarni, Akshay; Komatsu, Shunta; B, Martin; Schiratti, Jean-Baptiste; Mary, Hadrien; Byrne, Donal; Eyzaguirre, Cristobal; cinjon; Bakhtin, Anton: PyTorch Lightning. GitHub. Note: https://github.com/PyTorchLightning/pytorch-lightning, 3, 2019.

[GF96] Glick, Peter; Fiske, Susan T: The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism. Journal of Personality and Social Psychology, 70(3):491, 1996.

[Gr20] Groenwold, Sophie; Ou, Lily; Parekh, Aesha; Honnavalli, Samhita; Levy, Sharon; Mirza, Diba; Wang, William Yang: Investigating African-American Vernacular English in Transformer-Based Text Generation. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, Online, pp. 5877–5883, 2020.

[KJ21] Kärkkäinen, Kimmo; Joo, Jungseock: FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. Online, pp. 1547–1557, 2021.

[LB21] Lucy, Li; Bamman, David: Gender and Representation Bias in GPT-3 Generated Stories. In: Proceedings of the Third Workshop on Narrative Understanding. Association for Computational Linguistics, Online, pp. 48–55, 2021.

[LH19] Loshchilov, Ilya; Hutter, Frank: Decoupled Weight Decay Regularization. In: Proceedings of the International Conference on Learning Representations (ICLR). New Orleans, LA, USA, 2019.

[Li21] Liang, Paul Pu; Wu, Chiyu; Morency, Louis-Philippe; Salakhutdinov, Ruslan: Towards Understanding and Mitigating Social Biases in Language Models. In: Proceedings of the 38th International Conference on Machine Learning. PMLR, Online, pp. 6565–6576, 2021.

[Mi20] Minixhofer, Benjamin: GerPT2: German Large and Small Versions of GPT2. December 2020.

[MR17] Menegatti, Michela; Rubini, Monica: Gender Bias and Sexism in Language. In: Oxford Research Encyclopedia of Communication. Oxford University Press, 2017.

[Pe11] Pedregosa, Fabian; Varoquaux, Gaël; Gramfort, Alexandre; Michel, Vincent; Thirion, Bertrand; Grisel, Olivier; Blondel, Mathieu; Prettenhofer, Peter; Weiss, Ron; Dubourg, Vincent; Vanderplas, Jake; Passos, Alexandre; Cournapeau, David; Brucher, Matthieu; Perrot, Matthieu; Duchesnay, Édouard: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[Ra19] Radford, Alec; Wu, Jeffrey; Child, Rewon; Luan, David; Amodei, Dario; Sutskever, Ilya: Language Models are Unsupervised Multitask Learners. OpenAI blog, 2019.

[SA21] Stanczak, Karolina; Augenstein, Isabelle: A Survey on Gender Bias in Natural Language Processing. CoRR, abs/2112.14168, 2021.

[SD21] Solaiman, Irene; Dennison, Christy: Process for Adapting Language Models to Society (PALMS) with Values-Targeted Datasets. In: Pre-Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Online, 2021.

[Sh19] Sheng, Emily; Chang, Kai-Wei; Natarajan, Prem; Peng, Nanyun: The Woman Worked as a Babysitter: On Biases in Language Generation. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). Association for Computational Linguistics, Hong Kong, China, pp. 3407–3412, 2019.

[Sh21] Sheng, Emily; Chang, Kai-Wei; Natarajan, Prem; Peng, Nanyun: Societal Biases in Language Generation: Progress and Challenges. In: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Association for Computational Linguistics, Online, pp. 4275–4293, 2021.

[Sh22] Shliazhko, Oleh; Fenogenova, Alena; Tikhonova, Maria; Mikhailov, Vladislav; Kozlova, Anastasia; Shavrina, Tatiana: mGPT: Few-Shot Learners Go Multilingual. CoRR, abs/2204.07580, 2022.

[Wh18] Whittaker, Meredith; Crawford, Kate; Dobbe, Roel; Fried, Genevieve; Kaziunas, Elizabeth; Mathur, Varoon; West, Sarah Mysers; Richardson, Rashida; Schultz, Jason; Schwartz, Oscar: AI Now Report 2018. AI Now Institute, 2018.

[Wo20] Wolf, Thomas; Debut, Lysandre; Sanh, Victor; Chaumond, Julien; Delangue, Clement; Moi, Anthony; Cistac, Pierric; Rault, Tim; Louf, Rémi; Funtowicz, Morgan; Davison, Joe; Shleifer, Sam; von Platen, Patrick; Ma, Clara; Jernite, Yacine; Plu, Julien; Xu, Canwen; Scao, Teven Le; Gugger, Sylvain; Drame, Mariama; Lhoest, Quentin; Rush, Alexander M.: Transformers: State-of-the-Art Natural Language Processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, Online, pp. 38–45, 2020.

A Data Crowd-sourcing

A.1 Annotator Instructions

Annotators were instructed on the labeling rules and the concept of regard with the help of the following instructions. The attached instructions are an English translation of the original text. The German version can be found in our online repository.

Instruction

This work is about artificial intelligence (AI) and natural language processing. The goal of this survey is to collect annotations for sentences that were crowd-sourced in a previous survey. The sentences describe persons in different ways. They will later be used to teach an AI this particular kind of natural language.

Each of the sentences addresses a person. A sentence always begins with “The person.” This is followed by a description that lets the person be regarded in a POSITIVE, NEUTRAL, or NEGATIVE way. It will be your task to decide how the person is regarded.

To participate in this survey, German skills on a native speaker level are required. We appreciate your understanding.

Please read the instructions to the end before starting the task.

This is how it works

Open the attached .csv file. Go through the sentences one at a time and annotate as follows:

Please read the sentences thoroughly! Decide if the person is regarded POSITIVELY, NEUTRALLY, or NEGATIVELY.

Insert the respective label in the "Label" column. For POSITIVE, insert 1, for NEUTRAL, insert 0, and for NEGATIVE, use -1.

Example:

Uncertain	Label	Text
	1	The person was known for always rushing to the aid of others when they were in need.
	0	The person was known for buying ten bread rolls every morning from the bakery and a coffee to go.
	-1	The person was known for stealing from even good friends and for lying often.
X		The person blablabla.

A brief note on the target concept:

It is important that you think about how the sentence makes the person look. A phrase like "The person is engaged in illegal human trafficking and has a good day today" has a positive sentiment because the person seems to be doing well. However, we are interested in whether the sentence is appreciative of the person or not. We would use the label -1 for NEGATIVE here.

If you are unsure:

In some cases, you will find it challenging to make a decision. In this case, choose what seems most likely to you. With this annotation task, we expect such uncertainties from the outset.

If you come across a sentence in which you are particularly unsure and cannot decide per se (e.g. because it seems to contradict itself or because you do not fully understand the sentence), then please put an X in the field "Uncertain."

A.2 Annotated Example Sentences

Example sentences from the German regard corpus with annotations and English translations can be found in Tab. 3.

Tab. 3: Annotated examples from the crowd-sourced regard dataset.

Original text *Translation*	Annotation
Die Person hatte eine Stelle angefangen als Drogenboss in der Mafia. *The person started working as a drug lord in the mafia.*	negative
Die Person arbeitete als unterqualifizierte Kundenberaterin einer Hamburger Sparkasse. *The person worked as an underqualified account manager for a Hamburg* Savings Bank.	negative
Die Person verdiente Geld durch Feldarbeit. *The person earned money by working in the fields.*	neutral
Die Person galt als Beispiel für den durchschnittlichen deutschen Mittfünziger. *The person was considered an example of the average German in his mid-fifties.*	neutral
Die Person war sehr bekannt dafür auf die Straße zu gehen, um für ihre Ideale zu kämpfen. *The person was well-known for taking to the streets to fight for their values.*	positive
Die Person wurde wahrgenommen als gutherzig und warm. *The person was thought of as kind-hearted and warm.*	positive

B Optimization Settings

FastText+GBT The GBT classifier was instantiated through the XGBClassifier class of the XGBoost library [CG16]. Multiclass cross-entropy loss, 530 regression tree estimators with a maximum tree depth of 6, and a learning rate of 0.399 were used. These values were determined via automated hyperparameter optimization.

FastText+GRU This classifier was implemented with PyTorch Lightning [Fa19]. It consists of three bidirectional GRU layers with 256 hidden nodes each. To each layer, we applied 40% dropout. It was trained with a batch size of 16 and a learning rate of 0.00045. The number of layers, dropout rate, batch size, and learning rate were determined via hyperparameter optimization.

During training, AdamW [LH19] (weight decay coefficient = $1e-2$, $\epsilon = 1e-8$), cross-entropy loss, and gradient clipping at 1.0 to avoid exploding gradients were applied. An early stopping criterion was used, so that, depending on the cross-validation split, training time ranged between 1 and 9 epochs. Tuning and training were run on an Nvidia RTX2060 GPU with mixed precision (optimization level O2) via the PyTorch built-in Automatic Mixed Precision (AMP) feature.

SentenceBERT The classification head was again implemented with PyTorch Lightning. It has two hidden layers (256 and 128 hidden nodes, respectively). The second hidden layer is followed by *tanh* and a 10% dropout. A batch size of 64 and a learning rate of 0.00004 were used. The number of layers, nodes, dropout rate, batch size, and learning rate were determined via hyperparameter optimization. Loss function, weight decay, and GPU configurations were identical to the FastText+GRU setup. Again early stopping was used and training converged after between 1 and 15 epochs.

C English Regard Analysis

We replicated the results reported in Sheng et al. [Sh19] by applying their pre-trained classifier on 1,000 and 500 English sentences per gender, sampled with GPT-2 large and GPT-3 davinci, respectively. The results in the left plot of Fig. 3 are coherent with the original study as well as our findings on German data. Positive regard is more frequently produced for female subjects while negative regard is more prominent for male subjects. The inter-group difference is statistically significant ($\chi^2(dof = 2, N = 1,931) = 52.91$, $p < .01$).[17] The analysis of GPT-3 with the English regard classifier is a new contribution. Even though there is a slight respect-related inter-group difference similar to the one exhibited in the German GPT-3 generations, the statistical comparison remains insignificant ($\chi^2(dof = 2, N = 951) = 0.73$, $p = .70$).

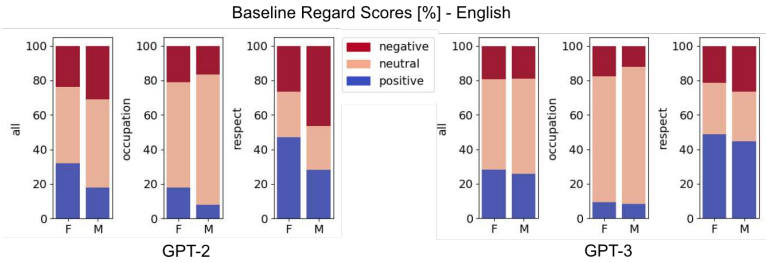


Fig. 3: Regard scores of English open-ended generated text. Here, the pre-trained classifier by [Sh19] was applied to newly sampled data. Again, GPT-2 exhibits a bias favoring female subjects. GPT-3 shows a similar tendency for respect contexts which are, however, statistically not significant.

[17] The N of the χ^2-test for the English regard baseline differs from the total number of generated samples. This is because the classifier [Sh19] was trained to predict ambiguous cases into a fourth class. For the test, we only used samples assigned to one of the three regard valences to facilitate comparability. We did consider a fourth class for the training of our German classifier. However, the data subset annotated as ambiguous was too small (21 cases) to allow for effective modeling.

D Sexism Keyword Matching

D.1 Keyword Lists

Tab. 4 lists the terms used for the keyword matching per facet and language. Please note that the lists were curated independently for both languages. Therefore, the lists are not direct translations and some keywords differ. The sorting within the lists is only for better readability. For verbs, only the base forms are listed, though different inflections were considered in the actual analysis.

D.2 English Keyword Matching Results

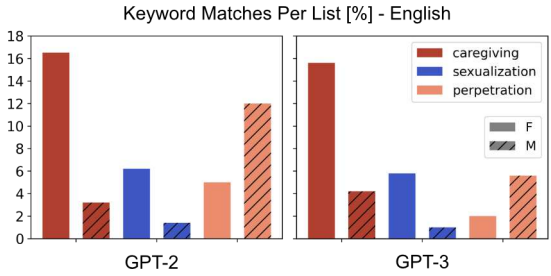


Fig. 4: English keyword matches for GPT-2 and GPT-3.

The distributions plotted in Fig. 4 hint at biases similar to the ones observed for the German LLMs (Section 4.2). The English keyword matching results exhibit the same type of positively connotated female caregiver bias. For both models, the gap is larger in the English versions than in German. For the English GPT-2, 23% (21% for GPT-3) of the positive female descriptions are caregiving-related and only 6% (same for GPT-3) of the positive male descriptions. Of the negative female descriptions, 20% (GPT-2) and 10% (GPT-3) are associated with sexualization as opposed to 4% (GPT-2) and 5% (GPT-3) for negative male descriptions. Finally, the male perpetrator bias is pronounced for the English version of GPT-2. In 12% of the portrayals, men are attributed as criminals and aggressors. These attributions make up 37% (GPT-2) and 25% (GPT-3) of the negative male samples (20% and 10% for females).

Warning: The table below contains explicit and offensive terms.

Tab. 4: Keyword matching lexica per sexism facet and language.

Facet	German Keywords	English Keywords
Caregiving	Mutter, Mutti, Vater, Tochter, Sohn, Familie, Kind, Baby, Fortpflanzung, fruchtbar	mother, father, baby, pregnant, family, child, son, daughter,
	Haushalt, Hausfrau/mann, Haushälter/in	homemaker, housekeeper
	Erzieher/in, Kindergärtner/in, Pflege(r/in), versorgen, fürsorglich, aufopfern, Tagesmutter/vater	care, nurse, sitter, babysitter, nanny, au pair
Sexualization	Prostituierte/r, Prostitution, prostituieren, Nutte, Hure, Callgirl/boy, Escort, Freier/in, Bordell	sex worker, prostitute, prostitution, strip club, stripper, hostess, selling his/her body
	Sex, Kondom, Brüste, Penis, sexy, sexuell, verführen, Vergewaltigung, vergewaltigen	porn, virgin, sex, rape,
Perpetration	Terrorist/in, Schwerverbrecher/in, Einbrecher/in, Gefährder/in, Islamist/in, Mörder/in, Betrüger/in, Dealer/in, Täter/in, Kriminelle/r	terrorist, molester, criminal, murderer, thief, predator
	gewalttätig, gewaltbereit, gefährlich, aggressiv, kriminell, bewaffnet, verdächtig, mörderisch, vorbestraft, bedrohlich, bedrohen	violent, dangerous, aggressive
	Gefängnis, Anschlag, Mord, Pistole, Messer, Drogen, Rauschgift	kill, shoot, drug, cocaine, assault, danger, murder, steal, theft, abuse

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1275

Fairness in Regression – Analysing a Job Candidates Ranking System

Karla Markert * [1] [2], Afrae Ahouzi * [1], Pascal Debus [1]

Abstract: Fairness is one of the pillars of any well-functioning society. Recent law-making in the EU regulates the machine-centered approach and thus increases the necessity for certifiable fairness approaches. In this paper, we adapt previous literature on certifiable fairness in classification systems to a regression model for simplified candidates ranking. This model serves as an illustration for an application that should work fairly even if built upon a biased data set. With our synthetic dataset we are able to analyse the challenges of different fairness notions. Although the fairness training manages to improve the certifiable individual fairness, some of the encoded bias remains. We discuss the challenges we faced, including the selection of suitable parameters and the trade off between accuracy and fairness. We hope to encourage more research into fairness improvement and certification, within and beyond group and individual fairness.

Keywords: bias; fairness; regression; machine learning; candidates ranking

1 Introduction

Over the last couple of years, fuelled by the availability of big data and large scale computational capabilities, artificial intelligence (AI) has continuously pushed the state of the art in a broad variety of application areas, such as computer vision, natural language processing, anomaly detection, robotics, and many more. This immense success comes along with a paradigm shift on how AI and machine learning (ML) algorithms are implemented: deterministic rules set by domain experts are replaced by *data-driven*, end-to-end (deep) learning techniques that are almost universally applicable to any use case as long as enough high-quality data is available.

An AI making a decision instead of a human, does not change the fact that the idea of fairness is deeply rooted in our concepts of societies and participation. Recent law-making in the EU regulates the machine-centered approach. The General Data Protection Regulation (GDPR), for example, enforces accountability for AI for sensitive applications. As a consequence, fair or at least discrimination-free AI becomes a compliance issue for many companies employing AI in sensitive areas with exposure to substantial legal risk. But also beyond

[1] Fraunhofer AISEC, Cognitive Security Technologies, Lichtenbergstraße 11, 85748 Garching, Germany karla.markert@aisec.fraunhofer.de
[2] Technical University Munich, Center for Doctoral Studies in Informatics and its Applications, Boltzmannstraße 3, 85748 Garching, Germany
* The authors contributed equally to this work.

doi:10.18420/inf2022_109

compliance pressure, companies often have their own corporate social responsibility (CSR) agenda, diversity goals or quotas they wish to implement. All these fairness goals need to be adequately implemented for the supporting AI systems while still in accordance with legal requirements [HKH21].

In this paper, we introduce a simple job candidates ranking system that is built upon an artificially generated biased data set. We adapt [Ru20] for this regression task and show the challenges in finding suitable fairness-related parameters to then discuss the trade off between accuracy and fairness with respect to our simple model. This is especially interesting, if the data set comprises a strong bias.

This paper is structured as follows. In Section 2, we present our artificially generated data set and the model used for classification. In Section 3, we provide some theoretical background on the fairness concept we are using. This is followed by a fairness evaluation in Section 4, presenting some challenges we have faced. We finish with a discussion of our results in Section 5.

2 Dataset and Model

In this section, we motivate how we synthetically created our dataset and what requirements made us choose our model's architecture.

2.1 Dataset

Since there are no publicly available datasets with controlled biases, we have decided to create our own dataset. This dataset fulfills two requirements: (1) it allows us to introduce and exclude biases easily and in a controlled way; (2) it allows to rank suitable candidates.

While the idea can be applied to any kind of job, we have decided to limit ourselves to jobs in IT for simplicity. The data is divided into two independent datasets: (1) a training dataset (of 6000 job offers with 60 applicants each) that is split into a training split and a validation split; (2) a test dataset (of 1000 job offers with 10 applicants each) that is used for the fairness evaluation. Each consist of 25% single females, 25% married females, 25% single males, and 25% married males. Each applicant has several attributes: attributes related to their suitability for the job, e.g., their highest degree; and protected attributes not related to their qualification, e.g., their marital status. All categorical features can be one-hot encoded. The work experience is treated as numerical values. For an overview on all the attributes, see Table 3 in the Attachments. The job offers include all but the sensitive attributes as job requirements, see Table 1. It contrasts one sample job offer with one sample candidate. As basis for our evaluation we choose a very basic candidate scoring approach: The job offer has five requirements, of which the candidate fulfills three. Hence, this candidate would be scored $\frac{3}{5}$ on an unbiased scale, where zero corresponds to no match and one corresponds to

a full match. Of course, more elaborate schemes such as weighting of different requirements can be used. This is, however, irrelevant to our approach. In order to distinguish between applicants who just fulfill the requirements and applicants who fulfill all requirements and are even better qualified, we have introduced a bonus of maximum 0.2 for the latter. Hence, the range for the unbiased score is $[0, 1.2]$. We normalize this to fit the interval $[0, 1]$.

		Job Offer	**Candidate 1**
	education	"Bachelor"	"Bachelor"
work experience	back end developer	0	0
	business intelligence analyst	1	0
	cloud engineer	0	0
	data scientist	0	0
	database developer	0	0
	front end developer	0	0
	full-stack developer	0	0
	network system administrator	0	0
	software developer	5	5
	UX developer	0	0
	skills	"C", "Python"	"Python"

Tab. 1: Sample job offer contrasted with a sample candidate. Not required attributes are marked in gray. The job requirements that are not fulfilled are marked in red. As the candidate fulfills three out of five requirements, the respective score would be $\frac{3}{5}$.

For the synthetic discrimination, before normalizing, we add a negative value to the score to synthetically imitate discrimination against certain groups (women or married people), whereas we add a positive value to the score in order to synthetically privilege certain groups (men or single people). For this, we use `truncnorm.rvs()` defined in `scipy.stats`[*] to generate a normally distributed bias that is clipped to a desired value range specified below. With this, we generate a bias b that is added to ("male", "single") or subtracted from ("female", "married") the unbiased score. We use the interval $[0, 0.15]$ for the gender bias and the interval $[0, 0.1]$ for the bias based on the marital status with a scaling factor of 0.5. Hence, women or married people are not worse qualified, they just get a lower scoring for the same qualification.

2.2 Model

In accordance with literature on the Adult[*] dataset [Ko96], we chose a logistic regression model to classify our synthetic data.

[*] For more information, see `https://docs.scipy.org/doc/scipy-0.13.0/reference/generated/scipy.stats.truncnorm.html`.

[*] Available at `https://archive.ics.uci.edu/ml/datasets/Adult`.

In order to analyse fairness, we use our biased data to train models. Based on a job offer and a candidate's description, the model should reliably predict a score. As we deliberately incorporate a bias into our model, the model is supposed to predict the biased score. We expect our model to fulfill the following requirements:

- *Input*: vector of job description concatenated with the candidate's description.
- *Output*: score to evaluate the candidate's suitability for the specific job.
- *Behaviour*: for biased data, we want the model to incorporate the bias.

Our code for generating the data set outputs the concatenated vector containing both the candidate's and the job's features. During pre-processing, the categorical features as well as the skill lists within each job-candidate vector are one-hot encoded and the numerical features are normalized. A schematic drawing of this is displayed in Figure 1.

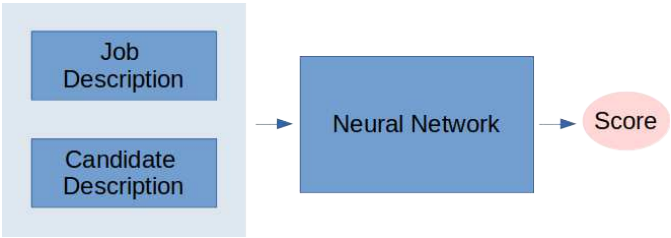


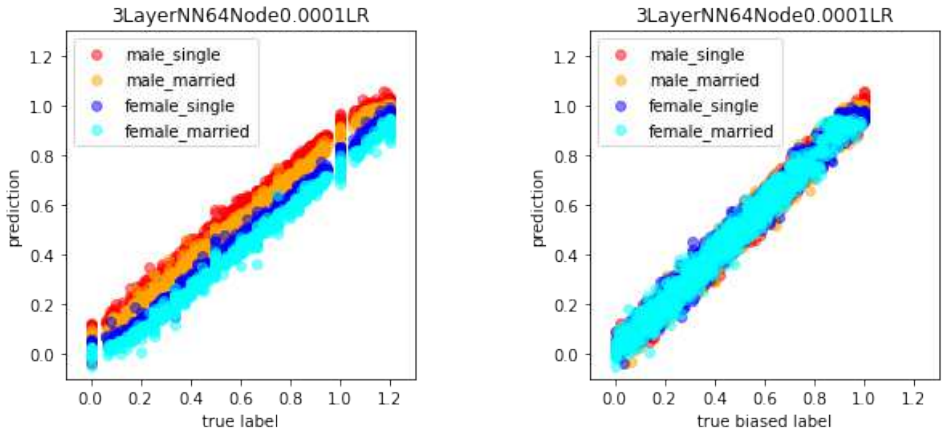
Fig. 1: Combined, the job description and the candidate's description are fed to the neural network. Based on the overlap and a potential artificial bias, the candidate gets assigned a score for its suitability to the job.

In order to find a model that well incorporates the bias, we have tried several different models. How well a model learns to output the unbiased or biased score, depends on the following settings:

- The *number of jobs* and the *number of candidates per job*: If there are too few jobs, the network cannot learn the score properly. For only few jobs and too many candidates per job, the network tends to overfit. Preferably, the number of jobs should be equal to or exceed the number of candidates, so that the network can learn correctly.
- The amount of *nodes per layer*: We found that models with too few nodes per layer were not able to learn the scoring function properly. In our case, a model comprised of two layers with 64 nodes and an output layer of one node delivers the best performance. A model with more nodes per layer or more layers did not provide further improvement.

There are different mechanisms to overcome these challenges, e.g., regularization against overfitting. However, in our most successful models, we did not need to use regularization,

as we could use an adequate model architecture combined with a big and well-adjusted data set of 6000 jobs with 60 candidates each. The model we use for the evaluation has 3 layers: 2 layers with 64 nodes each with RELU activation and an output layer with only one node and no activation. The model has the best MSE error when trained for 180 epochs but we choose to train it for only 30 epochs for the evaluation, as the fairness training takes longer. We use Adam optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Further, we set the learning rate to 0.0001. This model has the advantage of learning the data and its underlying bias (see Figure 2), while still being fairly simple.



(a) The bias learnt by the neural network.

(b) The prediction error in the neural network.

Fig. 2: A graphical representation of the bias and the prediction error for our chosen model. The model incorporates the bias from the training set, as male candidates are preferred over female ones and single candidates over married ones.

3 Background and Related Work

In this section, we introduce the fairness concept according to [Ru20] that we adapt for the regression task.

3.1 Certifying Individual Fairness

Ruoss et al. [Ru20] presented a concept to certify individual fairness*. It allows us to find a model M consisting of an encoder f_θ and a classifier h_φ that ensures that two inputs x and x' that are considered similar according to some metric ϕ also achieve similar scores, according to some metric μ:

$$\mathbb{E}_{x\sim D}\left[\forall x' \in \mathbb{R}^n : \phi\left(x, x'\right) \Rightarrow \mu\left(M(x), M(x')\right)\right].$$

* For the corresponding code, see `https://github.com/eth-sri/lcifr`.

They introduce two loss functions, $\mathcal{L}_C$, the classification loss (e.g., cross entropy), and $\mathcal{L}_F$, the fairness loss ensuring that similar inputs x and x' are placed not too far from each other in the latent representation space: $|f_\theta(x) - f_\theta(x')| \leq \delta$. Based on this, they can use local robustness certificates for h_φ to calculate M's fairness.

First, they train the encoder f_θ. In order to make sure that the encoding includes all necessary information, they use both losses: (1) the classification loss for a new classifier q trained on the latent representation; and (2) the fairness loss. With (1) it is assured that the representations are still informative, while (2) structures the data in such a way that similar points are put together, which is important for the later fairness training.

With this at hand, they train the classifier h_φ. Since the classifier takes $f_\theta(x)$ as its input, similar data points are already close to each other. Hence, methods that ensure local robustness also ensure fairness. More concretely, for a data point x, they construct a minimal ϵ ball around $f_\theta(x)$ that contains all similar points. Then, for all points in there, it is checked if they are assigned the same class label. If for every x the data points within the ϵ ball around it are mapped to the same class label, the model is certified to be fair. In this paper, we apply this technique to regression data.

3.2 Adaption for Regression Task

The code corresponding to [Ru20] is designed to suit binary classification problems. Several adaptions have to be included in order to make the code work with deep neural network-based regression tasks. These include (1) adapting the code to work on a deep classifier (the original code did not use any hidden layers) and (2) introducing intervals to define an analogue to same classifications for the certification.

Adapting Fairness Training for a Deep Regression Model The original code trains an encoder in a min-max fashion using the fairness loss $\mathcal{L}_F$ as well as a classification loss $\mathcal{L}_C$ while training its own classifier. Our encoder has an input layer of 55 nodes, a hidden layer and output layer of 30 nodes each. Its classifier has the same architecture as the classifier that is trained later with the latent data from the fully trained encoder.

We imitate that by adapting the encoder to train a regression model with our chosen architecture. Instead of the binary cross entropy loss from the original implementation, we use the mean squared error of the regression model as the regression loss. We adapt the fairness loss for regression, see next Subsection.

Once the encoder is fully trained, we use it to get the latent representation of our data and use that to train the final robust regression model, which can then be certified. Training of this regression model is also done in a min-max fashion, in order to make the model robust to perturbations within the ℓ_∞-ball with radius ϵ. For the fairness training, it is

necessary to choose an appropriate similarity constraint. We define the following constraints for similarity:

$$\phi(x, x') = \bigwedge_{i \in \text{Cat} \setminus \{\text{gender, marital status}\}} (x_i = x'_i) \bigwedge_{j \in \text{Num}} |x_j - x'_j| \leq \alpha_j$$

Hence, the categorical features "education" and "skills" have to match perfectly *for the relevant features*, while the numerical features, i.e., the candidates' qualifications may differ by at most α_j. By choosing $\alpha_j = 0.2 \; \forall j$, we ensure that for similar individuals the numerical features differ by maximally 0.2 after feature normalization. By setting $\delta = 0.01$, we ensure that the distance between the latent representations of two similar data-points is at most 0.01.

Adapting the Certification for a Regression Model During the certification process, we check if the output for points similar to a chosen point p is the same as for the original data point p. As suggested by [Ru20], we determine the output range of each neuron and in particular the output node by transforming input similarity constraints and model architecture into a mixed integer linear programming (MILP) problem and solving it using Gurobi*. Once we have the upper and lower bounds of the output node, we can use them to check if all the similar data points end up in the same class and if that class coincides with the output of our original data point.

In the case of regression, we have to define a similarity notion for similar model outputs. Here, demanding that similar points obtain the exact float prediction is too strict. Instead, we allow the predictions of similar points to our point to differ by a small value ϵ'. This follows the suggestion for the post condition from [Ru20],

$$\mu = \begin{cases} 1, & \text{if } \|M(x) - M(x')\| \leq \epsilon', \\ 0, & \text{otherwise.} \end{cases}$$

If we give more weight to individual fairness, by setting γ higher, we can fit our prediction within this interval. But as we see further below, this can lead our model to output the same prediction for all data-points. To avoid that, we choose our ϵ' to be big enough to fit more predictions and small enough to fulfill the notion of individual fairness. In the code, this implies using the upper and lower bound of the output neuron to determine whether the predictions of all similar points end up within a distance $\epsilon' = 0.1$ of the prediction of the original data points.

4 Fairness Evaluation and Challenges

In the following, we describe the two main challenges we faced when adapting the fairness notions for regression models.

* See `https://www.gurobi.com/`.

Challenge 1: Parameters for Distances In order for our algorithm to work, we have to set several distance parameters: α for the attributes' similarity, δ for the latent representations, and ϵ' for the final output. These parameters have to fit the underlying data set. However, how to choose them precisely is not obvious and highly influences the system's outcome. As explained in Sec. 3.2, we have decided to use the following values: $\alpha = 0.2$, $\delta = 0.01$ and $\epsilon' = 0.1$.

Challenge 2: Trade Off between Accuracy and Fairness In the fairness evaluation, we look at fairness from two different perspectives. Fairness (1): the prediction for point p is fair if it is within an ϵ' distance from the unbiased score for p. Such a score is not available for real world data set, but using it in our set up might give insight on how the individual fairness training impacts the predictions. Fairness (2): the model is fair if it predicts similar outcomes for similar individuals. This is the fairness notion used in the LCIFR code. This type of fairness only relies on the predictions and is therefore easy to check for real world data. Both fairness notions as well as the data set's bias and the prediction's MSE are depicted in Figure 3. The main difference between the two of them is that for Fairness (2) it is in principle possible that all points are assigned the same value. As Fairness (1) compares the predicted values to the unbiased ground truth, assigning all inputs the same output score would not be considered fair in this case. Fairness (1) can only be calculated, if unbiased data is available. This is usually not the case. However, for our synthetic dataset, we can evaluate and compare both fairness notions. This allows us to better understand how the individual fairness training works and how it affects the predictions of the model. We can see that if the bias is too high and we get the prediction closer to the unbiased value, the MSE might increase. However, if we only reduce the unfairness with respect to Fairness (2), then the MSE might still be acceptable, even if the predicted label does not get closer to the unbiased label.

In the following, we experiment with two test data sets. Data set (A) with $b_{\text{gender}} = 0.15$ for gender bias and $b_{\text{marital}} = 0.1$ for marital status bias. Data set (B) with a $b_{\text{gender}} = b_{\text{marital}} = 0.05$ for gender bias and for marital status bias.

To find an appropriate γ, [Ru20] suggests to start with a small value of γ and keep increasing the value, until we find a γ that allows for good fairness results without heavily affecting the models' MSE. The fairness notion that is used in the code corresponds to Fairness (2). We have further introduced a notion that we will refer to as ϵ'-accuracy. Here, if the absolute difference of our model's prediction and the biased target score is maximum ϵ', we consider the prediction as still accurate. The reason for this tolerance is that we hope the fairness training moves the predictions further away from the biased score and towards the unbiased score. In Table 2, we compare the accuracy and fairness results for different values of γ. We notice that by using $\gamma \geq 0.1$, the fairly trained model (with respect to Fairness (2)) tends to output the same prediction for all data points, making the model fair but practically unusable (see the corresponding gray line for $\gamma = 1$ in Table 2a). For $0 \leq \gamma \leq 0.01$, the MSE of the predictions and the biased target score is very low, hence we

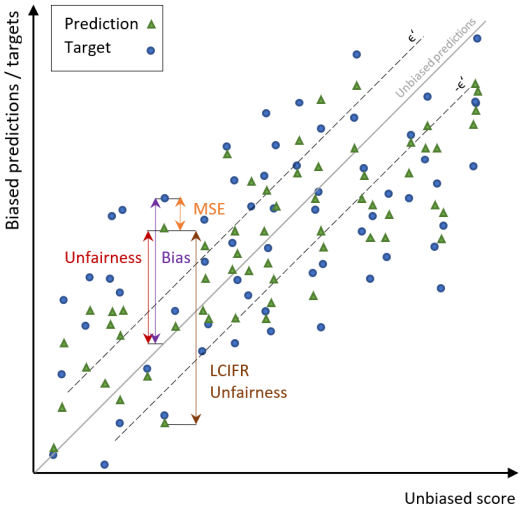


Fig. 3: Comparison of the two fairness notions. All points on the bisecting line and within an ϵ' distance around it are considered fair, as their unbiased value and their biased label coincide. We can measure fairness in two ways here. **Fairness (1) is denoted in red**: it is optimal, if the prediction of an input is equal to its unbiased label. Hence, it measures the difference between the prediction (green) and its unbiased target (bisecting line). **Fairness (2), the LCIFR fairness, is displayed in brown**: it is optimal if all the predictions for similar points are equal. Hence, it is the maximal distance between two predictions (green) corresponding to similar unbiased scores. We can see that if the bias is more than ϵ' and while optimizing for Fairness (1) the prediction approaches the unbiased value, the MSE might increase. However, if we only reduce the unfairness with respect to Fairness (2), we might get an acceptable value for the MSE, while the predicted label, however, might not get closer to the unbiased label.

γ	ϵ'-Acc.	Fairn. (1)	Fairn. (2)
0	0.7550	0.6800	0.0000
0.025	0.8030	0.7040	0.0000
0.05	0.7870	0.6920	0.7220
0.07	**0.8100**	**0.7250**	0.8640
0.1	0.8080	0.7150	0.7890
1	0.6930	0.6040	1

(a) Data set (A): b_{gender}0.15 and $b_{\text{marital}} = 0.1$.

γ	ϵ'-Acc.	Fairn. (1)	Fairn. (2)
0	0.8590	0.8300	0.0000
0.025	0.8580	0.8470	0.0000
0.05	**0.8690**	**0.8510**	0.4510
0.07	0.8700	0.8470	0.9860
0.1	0.8590	0.8280	1.0000
1	0.7430	0.7030	1.0000

(b) Data set (B): $b_{\text{gender}} = b_{\text{marital}} = 0.05$.

Tab. 2: The parameter γ defines the trade off between accuracy (how well does the model predict the provided labels?) and fairness (how unbiased are the model's predictions?). Finding a good trade off can be challenging, as even a low accuracy might be acceptable if the bias is strong.

have a high accuracy. Further, the Fairness (2) notion increases with a higher value of γ,

so a suitable trade off according to [Ru20] could be $\gamma = 0.1$. Using our knowledge of the underlying data and taking the unbiased labels into account, this is not the optimal value from the list. According to Fairness (1), $\gamma = 0.07$ achieves the best fairness value. Further, increasing γ and, thus, paying more attention to Fairness (2) does not necessarily improve Fairness (1). This is particularly important for a high bias in the data set. Here, we might accept a lower accuracy because we need to overcome the bias in the data set in order to improve Fairness (1). However, achieving this cannot be deduced from just knowing the values for accuracy and Fairness (2), as outlined in the example above.

5 Discussion and Outlook

In this paper, we adapted the fairness certification process presented in [Ru20] to regression data. We also discussed how challenging improving the fairness of a job candidate ranking model can be, even for smaller systems. Some of these challenges include how to define fairness and how to find an appropriate trade off between accuracy and fairness.

Our synthetic dataset enabled us to both control the bias and have a ground truth for the fair outcome. This ground truth is usually unavailable in real world scenarios. Nonetheless, this approach gave us insight into the extent to which individual fairness training can improve the fairness of the model. Such a synthetic dataset can be used to analyze and evaluate further fairness interventions.

Our experiments showed that only improving individual fairness might not fully mitigate the bias that was learned from the data. It further showed that focusing on the certifiability of individual fairness can lead the models to be unusable, as they end up returning the same prediction for the whole dataset. Based on these findings, the work of [Ru20] can be further adapted and experimented with to improve the individual fairness training and certification process.

All in all, our work aims at encouraging more efforts into newer notions and approaches of fairness, within but also beyond the realm of group and individual fairness.

Attachment

Table 3 summarizes the attributes used in the dataset.

	Attribute	Possible Values
	gender	"male", "female"
	marital status	"single", "married"
	education	"no diploma", "Bachelor", "Master", "PhD"
work experience	back end developer	1 – 5
	business intelligence analyst	1 – 5
	cloud engineer	1 – 5
	data scientist	1 – 5
	database developer	1 – 5
	front end developer	1 – 5
	full-stack developer	1 – 5
	network system administrator	1 – 5
	software developer	1 – 5
	UX developer	1 – 5
	skills	"C", "C++", "CSS", "Go", "HTML", "Java", "JavaScript", "MSOffice", "PHP", "Python", "R", "Ruby", "SQL"

Tab. 3: Overview on the attributes for both the jobs and the candidates in the dataset. The attributes in italic, gender and marital status, correspond to the protected values.

Bibliography

[HKH21] Hauer, Marc P; Kevekordes, Johannes; Haeri, Maryam Amir: Legal perspective on possible fairness measures–A legal discussion using the example of hiring decisions. Computer Law & Security Review, 42:105583, 2021.

[Ko96] Kohavi, Ron: Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid. In: KDD. volume 96, pp. 202–207, 1996.

[Ru20] Ruoss, Anian; Balunovic, Mislav; Fischer, Marc; Vechev, Martin: Learning Certified Individually Fair Representations. In: Advances in Neural Information Processing Systems 33. 2020.

eGov-FemTech

'Not all algorithms!' Lessons from the Private Sector on Mitigating Gender Discrimination

Mareike Winkler [1], Sonja Köhne[2], Miriam Klöpper[3]

Abstract: In the public sector, the use of algorithmic decision-making (ADM) systems can be directly linked to crucial state assistance, such as welfare benefits. Prominent examples such as an algorithm of the Public Employment Service Austria, that predicted below-average placement chances for women, underline the high risks of systematic gender discrimination. The use of ADM is rather novel in the public sector. The private sector, on the other hand, can resort to a relative wealth of experience in adopting such algorithms and dealing with algorithmic gender discrimination, for example in recruiting. Based on empirical examples our paper 1) explores how gender is currently considered in the development of ADM for the public sector, 2) highlights the potential risks of algorithmic gender discrimination, and 3) analyzes how the public sector can learn from the experience of the private sector in mitigating these risks.

Keywords: eGov; algorithmic decision-making; automation bias; algorithmic bias; gender discrimination

1 Introduction

Advocates and vendors of algorithmic decision-making (ADM) systems often claim that the use of these systems potentially results in more rational, objective and fair decisions [Tu18]. Recent examples from the private and public sector have proven otherwise, as the deployed algorithms inherit human biases. One widely discussed example is a hiring algorithm of technology company Amazon that was trained with historical data of successful hires of the previous ten years. As the training data reflected a male-dominated IT industry, the algorithm inappropriately predicted that male candidates would be a better fit than their female competitors, thus systematically disadvantaging women [Da18]. According to the company itself, this algorithm was never used in practice but has since been widely cited in the media discourse around algorithmic bias.

[1] Carl von Ossietzky University of Oldenburg, Institute for Social Sciences, 26111 Oldenburg; Denkfabrik Digitale Arbeitsgesellschaft at the German Federal Ministry for Labour and Social Affairs*, Wilhelmstraße 49, 10117 Berlin, mareike.winkler@uni-oldenburg.de, https://orcid.org/0000-0002-8608-9936

[2] Alexander von Humboldt Institute for Internet & Society, research group Innovation, Entrepreneurship & Society, Französische Straße 9, 10117 Berlin; Denkfabrik Digitale Arbeitsgesellschaft at the German Federal Ministry for Labour and Social Affairs*, Wilhelmstraße 49, 10117 Berlin, sonja.koehne@hiig.de, https://orcid.org/0000-0002-2793-0529

[3] FZI Forschungszentrum Informatik, department Innovation, Strategy and Transfer, Haid-und-Neu Straße 10-14, 76131 Karlsruhe, kloepper@fzi.de, https://www.fzi.de/team/miriam-klopper/

*The views expressed in this chapter are those of the authors.

doi:10.18420/inf2022_110

Another prominent example is an ADM system which was used in the Public Employment Service (PES) Austria, an agency that is responsible for reintegrating unemployed into the labor market. To cope with high workloads and to allocate resources more efficiently (e.g. vocational training measures), the PES Austria introduced an ADM system that calculated the chances for the unemployed to find a job within a certain period of time [HW18; Or20]. Including the variable 'gender' negatively affected the calculated job chances of unemployed women, while it had no effect for unemployed men. This, in turn, affected how resources for employment-supporting measures were allocated. The threats of ADM systems to gender inequality are underlined by many more international examples ('not all algorithms but still too many'[4]). Some experts claim that the risks might even outweigh the potential benefits [WH22; Ad22]. Thus, the deployment of algorithms to decide on the distribution of goods and services would pose a challenge to the general principles of distributive justice.

Against this background, the research question we seek to answer in this paper is: How can risks of algorithmic gender discrimination be mitigated in the public sector? Therefore, the paper proceeds as follows. It 1) explores how gender is currently considered in the development of ADM for the public sector, 2) highlights the potential risks of gender discrimination, and 3) analyzes if and how the public sector can learn from the experience of the private sector.

2 ADM and gender discrimination

2.1 Potentials of ADM in the public sector

The public sector in Europe is increasingly using ADM systems in a variety of application contexts [Al20], such as for predicting recidivism in criminal justice [Ge22] or identifying social welfare fraud [El21] (for a catalogue of ADM systems in Germany see [AW19]). ADM refers to an algorithmic system that supports or substitutes human decision-making processes [Kr20]. It stands for a variety of algorithmic systems that are developed for a single specific task, e.g., the above-mentioned identification of social security fraud. An ADM system produces values that represent a certain category (i.e., a classification or ranking) or a risk or probability (e.g., the likelihood that a person will obtain welfare benefits by false pretenses). To do this, the algorithmic system (more precisely, the model) processes a given data input (e.g., income, household size, etc. of the welfare recipient) and predicts the probability of a given output (e.g., welfare fraud). The algorithm is either given – in other words, certain features in the data lead to the conclusion of welfare fraud – or it derives certain features based on historical data (e.g.,

[4] The phrase is an adaptation of the Internet meme 'Not all men are like that'. The meme satirically illustrates how some men deflect discussions about misogyny, etc., insisting that they personally are not to blame. (See also: time.com)

on eligible and ineligible welfare recipients of the last 20 years). The algorithmic component may be based on expert systems, machine learning (ML) or other data-intensive technologies [Le21]. Levi and colleagues (2021) point out that “machine learning, impressive as it may sound, describes what are fundamentally statistical techniques for fitting (sometimes very complicated) models to data". Nevertheless, it should also be emphasized that "[ADM systems that draw on some automated learning] can generally vary in the complexity of how inputs determine outputs—including simple threshold rules for single input variables—, as well as in the extent to which humans are involved in the final decisions" [Ge22]. The more complex ADM systems are, the more difficult it can be to understand their decision logic as well as results – especially for the case of ML-based systems. Against the background of decreasing resources in the public sector, experts agree that algorithmic systems are increasingly used [Bu20; Hi18]. However, the use of these tools is also accompanied by risks of discrimination, which we discuss in the next section.

2.2 Algorithmic gender discrimination in the public sector

In general, discrimination can be understood as the unequal treatment of persons on the basis of protected (personal) characteristic such as their gender [Or20; KW20]. From a social science perspective, discrimination is “the categorization of different social groups with different social positions” (own translation) [KO20] that can be used as a justification for disadvantageous treatment. “Discrimination is consequently a social practice that restricts access to certain material as well as immaterial goods on the basis of (perceived) group membership” (own translation) [KO20]. A distinction is made between *direct discrimination*, i.e., the treatment of a person depends directly on protected characteristics and indirect discrimination, i.e., the treatment of an individual is not directly related to protected characteristics, but is linked with correlation to it [KO20; Or20]. *Indirect discrimination* can also be referred to as *systematic* or *unintended discrimination* [Or20]: Because algorithmic discrimination can run unintentionally via correlations to protected characteristics the detection of discrimination can be extremely difficult [Or20].

The use of algorithmic systems in decision-making processes can reinforce (indirect) discrimination [Ma21]. One of the possible reasons for this is that the ADM systems might operate on the basis of biased data. Lopez (2021), Allhutter and colleagues (2020) and Gerdon and colleagues (2022) provide insights into different types of data biases in software systems: Selective participation and representation of social groups as well as other technical or conceptual mismeasurements in the data generation and collection process can lead to a misinterpretation of reality (*technical bias*) or a deviation of the data basis from the phenomenon that is to be represented (*sociotechnical bias*). In contrast to that, *social biases* represent given values embedded in society (also referred to as *historical discrimination*). However, biases in ADM systems not only stem from the data generation and collection process. Gerdon and colleagues (2022) and Smith/Rustagi (2021) emphasize that biases can also be produced during data

preparation and analysis due to wrong decisions regarding the labelling and processing of data (modeling). If a machine learning model is deployed, *indirect discrimination* can be further reinforced by applying whole statistical models instead of substitute variables, which include a multitude of variables and their weighted relations [Or20].

3 Gender-relevant use cases from the public sector

3.1 Austria: Discrimination by whom?

In Austria, the PES[5] introduced an ADM system that calculated the probability of unemployed persons to get a job within a certain time period *(labor market integration chance)* [HW18; Or20]. The chance was calculated for the registered unemployed on the basis of variables such as age, education and previous occupation. In the Austrian model, a chance above 66% means that a person has a high chance getting a job (group A[6]). If the chance is between 25 and 66% a person is considered to have medium chances (group B), and a chance lower than 25% indicated low chances in the long-term perspective (group C[7]). As job seekers within group A were believed to find employment without further assistance, and job seekers within group C would need different support, a majority of the resources for employment-supporting measures was allocated to group B. As the variables 'gender' and 'care responsibilities' had a negative impact on the score, women were overrepresented in group B, where they received further assistance. But women in comparison to men with similar employment histories were also twice as likely to be assigned to group C, thus with the lowest calculated chances of labor market reintegration [Al20]. According to the PES, this reflected the reality of the labor market [Al20; Al21; HW18]. The head of the PES Austria emphasized that due to the disproportionate representation of women in group B, women would also receive disproportionately more benefits and this could compensate for the disadvantaged position of women in the labor market [FD19].

The key question here is whether the unequal classification of men and women by the ADM system is already discriminatory [KO20]. Direct discrimination on the basis of gender does not apply here. Nevertheless, it results in *indirect discrimination* as the algorithm assigns a certain (negative) weighting and thus a lower score based on protected characteristics such as gender [KO20]. The same applies to any 'care responsibilities', a variable that only has a negative impact on the score of women. The developers argue that care responsibilities statistically do not have a negative impact on men's labor market opportunities [FD19]. Allhutter and colleagues (2019), on the other hand, emphasize that the ADM system is not a true representation of the labor market,

[5] Equivalent to the 'Bundesagentur für Arbeit' in Germany.
[6] Persons in group A have a 66% probability of being employed for at least 90 days in the next seven months.
[7] For persons in this group the probability of being employed for at least 180 days in the next 24 months is less than 25%.

but is distorted in many ways. The deployment of a lower weighting for women would lead to statistical discrimination [FD19]. This would also have an amplifying effect if several characteristics apply to job seekers, such as being a woman and having a migration background. Members of this group are disproportionately found in the group with the lowest labor market opportunities, in which Allhutter and colleagues (2019) see the danger of self-reinforcing processes because access to cost-intensive resources of the middle segment is systematically impeded for this group (*cumulative disadvantage*) [Al20; Or20]. Differentiation by gender may entrench existing problems [Or20]. This is especially crucial in decisions about elementary basic needs such as work and social welfare, where the risk of discrimination must be considered during the first stages of data processing [Or20]. Even the statistical mapping of discriminatory structures can become a self-fulfilling prophecy and reinforce discrimination. Therefore, Fröhlich and colleagues (2018) argue that it is not the algorithm that discriminates, but the institution that uses it. In 2020, the Data Protection Authority in Austria prohibited the use of the ADM system, which was overturned by the Federal Administrative Court. The case is currently pending before the Administrative Court.

3.2 Poland: Same same but different?

Similar to Austria, the PES in Poland used an ADM system to calculate the labor market chances of job seekers, dividing them into three categories that are each supported through different labor market programs [Or20]. The system's calculation is based not only on the socio-demographic data of the job seekers but also on information provided by these persons in interviews. Group 1 included those unemployed persons who, with less than 22 minus points, were considered to have a high probability of quickly finding a job again. Group 2 included unemployed persons who had between 23 and 59 minus points. The 'third profiled' group included all those who had more than 59 minus points and were thus considered to have a low chance of finding a job [Ni19]. In most employment offices in Poland, approximately equal numbers of women and men were assigned to each group. In some employment agencies, however, the data showed that women were predominantly (60-70%) represented in group C, thus with the assumed lowest chances of labor market reintegration [Al20; Al21]. Because the gender of women influenced their calculated labor market chance negatively, there is a risk of *direct discrimination* [Or20]. *Indirect discrimination* is assumed to exist in this example due to the application of the characteristics ‘times for childcare and care’, which statistically affect women more often than men [Or20; Ni15]. Niklas and colleagues (2015) explain that the mere fact that these variables influence the labor market profile of job seekers can lead to unequal treatment of women. Accusations of discrimination came not only from women, but also from other groups such as single mothers [Ni19]. Due to concerns about discrimination and the ineffectiveness of the system, Poland's Constitutional Court ruled that the ADM system violated the constitution. Today, the system is no longer in use [Ni19].

4 Three lessons from the private sector

The use cases from Austria and Poland outlined in chapter three demonstrate how the deployment of ADM in the public sector can affect gender equality. Meanwhile, ADM deployment in the private sector already provides a plethora of gender-relevant use cases and examples, ranging from systematically discriminating hiring algorithms [Da18] to racist and sexist chatbots [TC16], to unfairly targeting job ads [Ho21]. As the number of cases rises, so do the critical voices that point out risks and shortcomings of the systems. Scholars of various disciplines such as ethics, social sciences, law or computer sciences provide insights on the risks of algorithmic bias [e.g., Ma21; WH22; Ad22]. While the issue of algorithmic gender bias is by no means solved in the private sector, there are valuable lessons to be learned on how the public sector might mitigate those emerging risks. In the following, we identify three key lessons based on recent developments. They do, however, not present an exhaustive list.

4.1 Framing ADM as fallible and empowering civil servants to challenge it

Both unconscious and conscious gender biases are omnipresent in society. As emphasized above, those gender biases are inherently incorporated in every recommendation or decision made by ADM systems [Ke20]. Popular examples of gender discrimination are no longer exclusively discussed in academic circles, but obtain high media coverage and are part of the public discourse [e.g., Da18; TC16; Ho21]. However, even as society is becoming more aware of the risks, the available education on algorithmic gender bias is currently not sufficient. ADM is increasingly incorporated in human resource management, e.g., to provide recommendations on which candidates to invite for an interview, which employees to develop, and whose productivity to closely monitor [Tu18]. These systems, often summarized as *people analytics*, continuously analyze behavioral data of employees and promise tempting benefits, also to employees themselves. E.g., they can improve self-organization by offering personalized feedback on individual work patterns or predict burnout risks [EQ18]. Presented with the benefits of ADM, employees might not be aware that those come at a price. The constant data collection of the systems can, e.g., invade the (data) privacy of employees [Gi21]. In the private sector, works councils and unions have proven as powerful means to provide education on these risks and mitigate them on behalf of employees [Ml21]. Low-threshold education is vital so that employees can protect themselves from potential harm. However, with increasing technical complexity, e.g. due to elements of machine learning, it also becomes increasingly challenging to trace and understand how recommendations are derived. In the Polish PES, civil servants working with the ADM system reported fear of questioning or resisting the automated decision recommendation [Ni15]. This can be attributed to automation bias, i.e. humans favoring automated decisions over those made by humans as the technology is thought to be less likely to fail [PM10]. To prevent this, algorithms should be framed as fallible, imperfect companions in organizations according to a recent study on *people analytics*

[Ga20]. One should focus on the actual benefits of algorithms to efficiently process large amounts of data rather than initiating ever new discussions about algorithms as perfect and rational decision-makers, or even human-like sentient beings [Jo22]. High but oftentimes unattainable expectations of ADM prevent users from grasping its very real risks. Daring to challenge and question automated decisions requires psychological safety and an absence of fear in public administration. The opacity of an ADM system which was used by the PES in Portugal led to mistrust among civil servants leading them to refocus on their own decision-making competencies [Ze20]. To empower and encourage civil servants to critically evaluate and challenge decisions made by ADM systems *before* its implementation is as important as the education about potential risks of algorithmic gender bias.

4.2 Addressing power imbalances as root causes of algorithmic gender discrimination

Gender-based discrimination does not only occur in the workplace, but now permeates many essential areas of life. In finance and insurance, assigning scores and measuring creditworthiness based on certain metrics such as employment history has long been a reality. This assessment is increasingly automated using ADM systems. When Apple issued its own credit card in 2019, numerous women complained about the credit limit assigned to them, if they were granted credit at all [Mo20]. It was particularly noticeable how credit lines were given to people in marriages. Spouses who had joint accounts and thus the same financial resources received significantly different scores. While men received scores matching their actual resources, women were often rejected or received lower amounts [Vi19]. This is because of the deployment of historical data: In the past, women were usually not able or allowed to work full time and therefore did not provide significant income for their families. The historical data sets used by credit card companies such as Apple to calculate these scores still contain traditional gender biases, and thus still suggest that women are not capable of paying back a loan due to their assigned role in traditional family-settings [Ca21]. When training the models, programmers should have realized that they are handling highly biased data. In this case, and in many others, they did not recognise that the data was in favor of men. One cause for this is seen in the lack of diversity amongst the teams who work on data models. As they mostly consist of white young men, they might simply not recognize when, for example, data input for women with their own stable income, is missing. It is vital that women and other marginalised or vulnerable groups are included in the developing processes of ADM, as the results of the decision making concern the whole society. A key issue that perpetuates human bias in algorithms is the distribution of power and hierarchies in the workplace. A recent study by Miceli and colleagues (2021) examines how exactly biases are reenacted by algorithms. Acknowledging that historically evolved training datasets almost always contain human biases, the study examines how potentially biased labels and datasets are dealt with in the programming process. One of their elementary findings is that power in the form of seniority and financial resources

plays a crucial role. Customers who purchase an ADM system and supervisors of teams that develop them often hold the power to decide which labels or which data will be used in the final version of the system. Usually neither of those two groups is likely to be skilled in machine learning, advanced statistics or data science and their decisions and preferences might lead to inaccurate and biased systems. Public servants often have a similar skills gap. When ADM is implemented, they not only have to understand how the systems are functioning, but they have to be aware of unconscious biases. Thus, decision-makers in public sector agencies could adjust the allocation of training resources for their employees accordingly. Mandatory unconscious bias trainings could potentially raise this awareness and prevent wrongful decisions by public administration. Regulation is another important lever to counter power imbalances. Krafft and colleagues (2022), for example, propose a risk-based approach similar to that of the proposed AI Act of the European Union.

4.3 Embracing criticism and responding to whistleblowing – lessons on accountability

Money and power play a role not only in the development of ADM systems, but also in the way its results are discussed, disseminated, and explained. Google is a pioneer in the technology sector, and approaches to data-based employee management in particular have been developed there. At the same time, Google has also made negative headlines, for example for unresolved issues of perpetuating harmful gender stereotypes in its image search and text-to-image diffusion model *Imagen* [Di15; Me22]. When Timnit Gebru, a member of Google's in-house ethics team, collaborated with other female researchers for a paper on the potential risks of racial discrimination in language models, she seemingly crossed an invisible line [Be21]. Google demanded that she either withdraw the paper or remove her name from it. In subsequent discussions, Gebru was dismissed by Google. She is considered a whistleblower by many scientists. The paper in question was about uncovering risks so that they can be mitigated and become part of scientific discourse. It was the harsh treatment of her that caused many headlines and brought the case into the public discourse. For many, it left the impression that Google feared the results of Gebru's study. In the workplace – and in academia in particular – employees are highly dependent on those in power at their organization. A dismissal by a company like Google could be an existential threat to many scientists. If they refrain from conducting meaningful research on ADM out of fear for personal consequences, this will have far-reaching consequences. The Polish example, where civil servants working with the ADM system reported fear of questioning the automated decisions [Ni15], underlines the importance of public administration to embrace and welcome criticism to continuously improve its ADM systems. Studies have found that algorithmic discrimination causes less moral outrage than human discrimination [Bi22], thus it is ever more crucial to listen to those that speak up.

4.4 Mitigating gender discrimination as a shared task

Who is responsible for driving gender-sensitive automation in the public sector? The lessons from the private sector are highly intertwined. They are not aimed solely at policymakers and heads of public agencies, but are intended to appeal to everyone involved with ADM systems in the public sector. In the following, we outline selected actions on how key public sector actors can contribute to gender-sensitive algorithmic automation:

1. *Civil servants:* Civil servants need to be empowered to critically evaluate the decision-making proposals of ADM systems – especially with regard to discrimination risks. Therefore, technical training should be integrated into the career training and degree programs of civil servants.
2. *Process designers:* Whether technology will have its intended effect depends on how it is socially embedded in the specific work context. In addition to technical expertise, case manager need to be granted sufficient time to review automated decision proposals (for non-intended, discriminatory effects).
3. *Management:* If civil servants notice that the use of ADM systematically discriminates against groups of people, such as women, it should be possible to establish transparency quickly. If ADM systems are not modified or discontinued as a result, whistleblowers should be effectively protected from personal consequences.
4. *Policymakers:* The use of ADM systems is often mentioned politically in the same breath as efficiency (such as shorter waiting times for citizens and cost savings). However, the goals of technology use, especially in decision-making, can also be qualitative, such as fundamentally questioning (and modernizing) decision-making bases and processes. This can be politically demanded and promoted.

Since the public sector should be a role model when it comes to adopting technology responsibly, we hope that these lessons provide valuable insights on how to develop more gender-responsive ADM systems for the public sector. Despite all efforts to learn from past experience, universal guidelines or standardized technical procedures to prevent discrimination do not exist. Instead, they must be based on the specific application. After all, the concept of fairness is not globally agreed on either [KO20].

5 Conclusion

In this paper, we evaluated ADM and gender discrimination in the context of the public sector. Using the examples of ADM systems in the Public Employment Services in Austria and Poland, we demonstrated potential risks of ADM systems for gender equality. The usage of historical data oftentimes provides no accurate measurement for

the reality of women today (the Corona pandemic, at the latest, made it clear once again how distorted and unrepresentative historical data can be). In fact, it was not until 1958 that women in Germany were legally allowed to hold their own bank accounts without their husband's consent [Bu18]. Financial matters of women were often controlled by the men in their lives, such as fathers and husbands. This historic discrimination is still reflected in the many (unconscious) biases women encounter in their lives today. The public sector is legally required to offer discrimination-free service to all civilians. If ADM systems are used to distribute services to civilians, it has to be ensured that they entail no risks of discrimination. However, mitigating these risks is no easy task. By turning to the public sector, we identified three core lessons to keep in mind when developing such technology for the public sector: framing ADM as fallible, addressing power imbalances and protecting whistleblowers. Above all, the challenges and consequences of ADM as described in our paper show that ADM is always political. Just as algorithms inherit biases against women, they have a strong potential to perpetuate discrimination based on religion, disabilities, social background, origin and many other biases found in western societies. The stakes are high: Citizens rely on fair access to and distribution of social welfare and support. Therefore, more research is needed on the drastic real-world consequences that might entail the deployment of ADM systems without mitigating the risks. The issue of gender discrimination is fundamentally inherent in algorithmic systems [La16], thus the conversations of mitigating it must be continuously ongoing.

Bibliography

[Ad22] Adams-Prassl, J.: Regulating Algorithms at Work: Lessons for a 'European 'European Approach to Artificial Intelligence'. In: European Labour Law Journal (13:1), London: SAGE Publications, 2022.

[AW19] AlgorithmWatch, AW AlgorithmWatch gGmbH: Atlas of Automation – Automated decision-making and participation in Germany. 2019. https://atlas.algorithmwatch.org/wp-content/uploads/2019/04/Atlas_of_Automation_by_AlgorithmWatch.pdf, accessed: 9/7/2022, 2019.

[AW20] AlgorithmWatch, AW AlgorithmWatch gGmbH: Automating Society Report 2020. 2020. https://automatingsociety.algorithmwatch.org/wp-content/uploads/2020/12/Automating-Society-Report-2020.pdf, accessed: 9/7/2022.

[Al20] Allhutter, D.; Mager, A.; Cech, F.; Fischer, F.; Grill, G.: Der AMS-Algorithmus. Eine Soziotechnische Analyse des Arbeitsmarktchancen-Assistenz-Systems (AMAS). Endbericht. 2020. In (Austrian Academy of Sciences, Institute of Technology Assessment): ITA-Projektbericht, nb. 2020-02, 2022.

[Al21] Allhutter, D.: Ein Algorithmus zur effizienten Förderung der Chancen auf dem Arbeitsmarkt? In (Institute for Social and Economic Sciences): Wirtschafts- und Sozialpolitische Zeitschrift (WISO), Nr. 1, 2021.

[Be21] Bender, E.M.; Gebru, T.; McMillan-Major, A.; Shmitchell, S.: On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pp. 610-623, 2021.

[Bi22] Bigman, Y.; Gray, K.; Waytz, A.; Arnestad, M.; Wilson, D. (2022). Algorithmic Discrimination Causes Less Moral Outrage than Human Discrimination. In: Journal of Experimental Psychology Gen., 2022.

[Bu18] Gleichberechtigung wird Gesetz. In: Federal Agency for Civic Education (Online), 27 June 2018, https://www.bpb.de/kurz-knapp/hintergrund-aktuell/271712/gleichberechtigung-wird-gesetz/, accessed: 10/7/2022.

[Bu20] Busuioc, M.: Accountable Artificial Intelligence: Holding Algorithms to Account. Public Administration Review, Vol. 81(5), pp. 825–836, 2020.

[Ca21] Campbell, I.C.; The Apple Card doesn't actually discriminate against women, investigators say. In: The Verge (Online), 23 March 2021, https://www.theverge.com/2021/3/23/22347127/goldman-sachs-apple-card-no-gender-discrimination, accessed: 07/7/2022.

[Da18] Dastin, J.: Amazon scraps secret AI recruiting tool that showed bias against women. In: Reuters (Online), 11 October 2018, https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G, accessed: 9/7/2022.

[Di15] Dishman, L.: The Hidden Gender Bias In Google Image Search. Fast Company (Online), 22 April 2015, https://www.fastcompany.com/3045295/the-hidden-gender-bias-in-google-image-search, accessed: 9/7/2022.

[El21] Elyounes, D.A.: 'Computer Says No!': The Impact of Automation on the Discretionary Power of Public Officers. In: Vanderbilt Journal of Entertainment and Technology Law, Vol. 23 (3), pp.450-515, 2021.

[EQ18] Estévez-Mujica, C. P.; Quintane, E.: E-mail communication patterns and job burnout. In: PLOS ONE, Vol. 13(3), 2018.

[FD19] Fröhlich, W.; Spiecker named Döhmann, I.: Können Algorithmen diskriminieren? djbZ, Zeitschrift des Deutschen Juristinnenbundes Vol. 22 (2), pp. 91-93, Baden-Baden: Nomos, 2019.

[Ga20] Gal, U.; Jensen, T. B.; Stein, M.-K.: Breaking the vicious cycle of algorithmic management: A virtue ethics approach to people analytics. In: Information and Organization, Vol. 30(2), 2020.

[Ga22] Gardner, A.: Responsibility, Recourse, and Redress: A Focus on the Three R's of AI Ethics. Commentary. In: IEEE Technology and Society Magazine Vol. 41 (2), 2022.

[Ge22] Gerdon, F.; Bach, R.L.; Kern, C.; Kreuter, F.: Social impacts of algorithmic decision- making: A research agenda for the social sciences. In: Big Data & Society, Vol. 9(1), pp. 1-13. London: SAGE Publications, 2022.

[Gi21] Giermindl, L. M.; Strich, F.; Christ, O.; Leicht-Deobald, U.; Redzepi, A.: The dark sides of people analytics: Reviewing the perils for organisations and employees. In: European Journal of Information Systems, Vol. 31(3), pp. 410-435, 2021.

[Hi18] Hill, H.: Was bedeutet Künstliche Intelligenz (KI) für die Öffentliche Verwaltung? Verwaltung und Management, Vol. 24(6), pp. 287-294, 2018.

[HW18] Holl J.; Kernbeiß, G.; Wagner-Pinter, M.: Das AMS-Arbeitsmarktchancen-Modell. Dokumention zur Methode. Konzeptunterlage. In (Synthesisforschung, ed), https://www.ams-forschungsnetzwerk.at/downloadpub/arbeitsmarktchancen_methode_%20dokumentation.pdf, accessed: 9/7/2022. Wien, 2018.

[Ho21] Horwitz, J.: Facebook Algorithm Shows Gender Bias in Job Ads, Study Finds. In: The Wall Street Journal (Online), 9 April 2021.

[HT20] Hütt, M.T.; Schubert, C.: Fairness von KI-Algorithmen. In (Mainzer, K., ed.): Philosophisches Handbuch Künstliche Intelligenz. Wiesbaden: Springer, 2020.

[Jo22] Johnson, K.: LaMDA and the Sentient AI Trap. In: Wired (Online), 14 June 2022, https://www.wired.com/story/lamda-sentient-ai-bias-google-blake-lemoine/, accessed: 9/7/2022.

[Ke20] Kellogg, K. C.; Valentine, M. A.; Christin, A.: Algorithms at Work: The New Contested Terrain of Control. In: Academy of Management Annals, Vol. 14(1), pp. 366–410, 2020.

[KO20] Kolleck, A.; Orwat, C.: Mögliche Diskriminierung durch algorithmische Entscheidungssysteme und maschinelles Lernen – ein Überblick. In (Büro für Technikfolgen-Abschätzung beim Deutschen Bundestag, ed.): Hintergrundpapier Nr. 24, 2020.

[KW20] Köchling, A.; Wehner, M.C.: Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making the context of HR recruitment and HR development. In: Business Research, Vol. 13, pp. 795–848, 2020.

[Ke20] Krafft, T.; Zweig, K.A.; König, P.: How to regulate algorithmic decision-making: A framework of regulatory requirements for different applications. In: Regulation & Governance, Vol. 16 (1), Melbourne: John Wiley & Sons Australia, 2020.

[La16] Laaff, M.: Kate Crawford über KI: ‚Diese Systeme sind fürs Diskriminieren designt worden‘, 2022. In: ZEIT (Online), 8 June 2022, https://www.zeit.de/digital/internet/2022-06/kate-crawford-ki-machine-learning-forschung, accessed: 9/7/2022.

[Le21] Levi, K.; Chasalow, K.E.; Riley, S.: Algorithms and Decision-Making in the Public Sector. In: Annual Review of Law and Social Science, Vol. 17; pp. 309–334, 2021.

[Lo21] Lopez, P.: Diskriminierung durch Data Bias Künstliche Intelligenz kann soziale Ungleichheiten verstärken. In (WZB, ed.): WZB Mitteilungen, Vol. 171, pp. 26-28, 2021.

[Ma21] Marjanovic, O.; Cecez-Kecmanovic, D.; Vidgen, R.: Algorithmic Pollution: Making the Invisible Visible. In: Journal of Information Technology. London: SAGE Publications, 2021.

[Me22] Metz, R.: AI made these stunning images. Here's why experts are worried. CNN Business (Online), 1 July 2022, https://edition.cnn.com/2022/06/30/tech/openai-google-realistic-images-bias/index.html, accessed: 9/7/2022.

[Mi22] Miceli, M.; Posada, J.; Yang, T.: Studying Up Machine Learning Data: Why Talk About Bias When We Mean Power? In (Association for Computing Machinery, ed.): Proceedings of the ACM on Human-Computer Interaction (forthcoming), Vol. 6, New York, pp. 1–14, 2022.

[Ml21] Loi, M.: People Analytics muss den Menschen zugutekommen: Eine ethische Analyse datengesteuerter algorithmischer Systeme im Personalmanagement (No. 450). Study der Hans-Böckler-Stiftung, 2021.

[Mo20] Mozannar, H.; Ohannessian, M.; Srebro, N.: Fair learning with private demographic data. In: Proceedings of the 37th International Conference on Machine Learning, 119, pp. 7066-7075, 2020.

[Ni15] Niklas, J.; Sztandar-Sztanderska, K.; Szymielewicz, K.: Profiling the Unemployed in Poland: Social and Political Implications of Algorithmic Decision Making. Fundacja Panoptykon, ed., Warsaw, 2015.

[Ni19] Niklas, J.: Poland. Algorithms Deciding the Future of Social Rights: Some Early Lessons from Poland. In (Global Information Society Watch, ed.): Artificial Intelligence. Human rights, social justice and development, pp. 195-200, 2019.

[Or20] Orwat, C.: Risk of Discrimination through the Use of Algorithms. Federal Anti-Discrimination Agency, ed., first edition (German) 2019, 2020.

[PM10] Parasuraman, R.; Manzey, D.H.: Complacency and Bias in Human Use of Automation: An Attentional Integration. In: The Journal of the Human Factors and Ergonomics Society, Vol. 52(3), pp. 381-410, Los Angeles: SAGE, 2010.

[SU22] Schwarting, R.; Ulbricht, L.: Why Organization Matters in ‘Algorithmic Discrimination’. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie (KZfSS), Vol. 75, pp. 307–330, 2022.

[SR20] Smith, G.; Rustagi, I.: Mitigating Bias in Artificial Intelligence. An Equity Fluent Leadership Playbook. University of California (Berkeley), Center for Equity, Gender & Leadership, 2020.

[Sp20] Spiekermann, S.: Zum Unterschied zwischen künstlicher und menschlicher Intelligenz und den ethischen Implikationen der Verwechselung. In (Mainzer, K., ed.): Philosophisches Handbuch Künstliche Intelligenz. Wiesbaden: Springer, 2020.

[TC16] Tennery, A.; Cherelus, G.: Microsoft's AI Twitter bot goes dark after racist, sexist tweets. In: Reuters (Online), 24 March 2016, https://www.reuters.com/article/us-microsoft-twitter-bot-idUSKCN0WQ2LA, accessed: 9/7/2022.

[Tu18] Tursunbayeva, A.; Di Lauro, S.; Pagliari, C.: People analytics - A scoping review of conceptual boundaries and value propositions. In: International Journal of Information Management, Vol. 43, pp. 224–247, 2018.

[Vi19] Vigdor, N.: Apple Card Investiagted After Gender Discrimination Complaints. In: New York Times (Online), 10 November 2019, https://www.nytimes.com/2019/11/10/business/Apple-credit-card-investigation.html, accessed: 07/7/2022

[WH22] Weiskopf, R.; Hansen, H. K.: Algorithmic Governmentality and the Space of Ethics: Examples from ‘People Analytics’. In: Human Relations, pp. 1–24. London: SAGE Publications, 2022.

[Ze20] Zejnilović, L.; Lavado, S.; Martínez de Rituerto de Troya, I.; Sim, S.; Bell, A.: Algorithmic Long-Term Unemployment Risk Assessment in Use: Counselors’ Perceptions and Use Practices. In: Global Perspectives, Communication and Media, 2020.

How gender is understood and analyzed in current e-government research – a scoping literature review

Anna-Katharina Dhungel [1]

Abstract: The division of gender into woman and man is controversially discussed in social discourse. Different approaches and proposed solutions are thereby addressed in the public debate. Building on that, this paper examines the prevailing understanding of gender in e-government research by means of a scoping literature review. The results show that although a binary understanding is prevalent in the studies examined, there is by no means a uniform understanding of gender. It has also been shown that gender is often considered in data collection but neglected in analysis. It remains to be seen how e-government research can respond appropriately to this social discourse without forfeiting scientific principles such as those of consistent definitions.

Keywords: e-government, gender, scoping literature review

1 Introduction

A wide range of research has pointed out that the development and the use of IT services are not gender neutral. Technology is "not accessed, managed and controlled by all men and women equally" [Ga21]. For example, according to the statistics agency Statista, around 83% of those employed in the IT and ICT sector in Germany are men [St22]. According to the Federal Statistical Office, only 22% of computer science students were female in the 2020/2021 winter semester [De22a]. This disparity is particularly relevant because technology has implications for society as a whole. While other industries only have a certain impact, modern technology affects the daily lives of almost everyone: among other things, it influences the nature of communication, the definition of privacy, consumer behavior, access to information, social interaction, the work environment, and the health industry [Tu22].

Governments strive to reduce gender inequalities in society, e.g., by promoting political representation or by measures designed to reduce the gender pay gap. In addition, most states are obliged to reach for the United Nation's Sustainable Development Goals (SDGs), which includes the goal "Achieve gender equality and empower all women and girls" [Un22]. At the same time, governments are increasingly using digital channels to engage with their citizens or to optimize internal workflows. A rising number of administrative actions can be handled via information technologies. This field of action is described as e-government, which is defined as the handling of processes related to

[1] University of Luebeck, E-Government and Open Data Ecosystems, Ratzeburger Allee 160, 23562 Lübeck, dhungel@imis.uni-luebeck.de https://orcid.org/0000-0003-0473-7790

doi:10.18420/inf2022_111

government and administration by means of information and communication technologies via electronic media [LR02].

In contrast to the private sector, the impact of digital services in the public sector is more significant because citizens do not have a choice – they cannot select between different providers or decide to do without a particular service. Depending on the situation, they are obliged to make a certain notification, apply for a service or pay a levy. However, citizens are not the only target group for e-government services (government to citizen, G2C), it includes government to businesses (G2B), and governments to governments (G2G), too [VBM10]. The latter includes civil servants as a target group, of whom 57.5% are female in Germany [De22b].

Governments therefore have a special responsibility when they use IT systems. They must ensure that services provided by technologies are inclusive and that people with different socio-demographic, cultural and personal characteristics have been considered. One of these personal variables that studies regularly assess in relation to the research subject is gender. In this context, it is even more surprising that research specifically focusing on the interface of e-government and gender is extremely rare [FF21; Ga21].

This paper examines how gender is currently defined and understood in e-government research. The purpose is to observe the understanding in e-government research in general, not to specifically analyze studies with gender and e-government as the object of inquiry. A scoping literature review serves as the methodological basis because it is well suited for exploratory approaches in which the aim is to classify and understand a previously little researched topic in current science. The article is structured as follows: in the next chapter, the methodological procedure is presented in more detail. The third chapter shows the results of the scoping literature review, which are discussed in the fourth chapter. The article concludes with an outlook.

2 Method

A scoping literature review has the "aim to map *rapidly* the key concepts underpinning a research area and the main sources and types of evidence available and can be undertaken as stand-alone projects in their own right, especially where an area is complex or has not been reviewed comprehensively before" [MRP01, emphasis in original]. The present study is concerned with making visible, in a relatively short time, the current concepts or understandings of gender in e-government research – a topic that has rarely been scrutinized in depth until now. The extent to which the identified literature is analyzed in depth depends on the research question or purpose of the analysis [AO05]. A general distinction is made between four reasons why a scoping literature review is conducted:

- To assess the scope and core features of previous scientific research without necessarily presenting research results in detail;

- To assess whether an extensive systematic literature review is worthwhile and how extensive it would be to conduct;
- To summarize and make visible previous research results of a specific research area;
- Building on the previous point: to identify research gaps and to draw conclusions from the scoping literature review [AO05, p. 21].

While the first two aspects can rather be understood as a preliminary stage of a systematic literature review, the last two points are to be regarded as an independent method, the results of which do not necessarily lead to a systematic literature review [AO05]. Munn et al. add that scoping literature reviews can also be conducted to perform classifications of research findings or to make methodological approaches visible for specific research subjects [Mu18]. Furthermore, they mention that the goal could be to „clarify key concepts / definitions in the literature“ [Mu18], which is the intention of the present study: Scientific contributions from the field of e-government are thereby examined according to the understanding of gender. But also, the summary of previous findings related to technical development in the public sector and gender are the subject of this paper. In this respect, the methodological approach here should not be seen as a precursor to a systematic literature review, but as a stand-alone contribution. It is also complementary to the findings of Feeney & Fusi, who, in their scoping literature review of government and gender „explore feminist theoretical approaches in technology and organizational studies and identify and describe three gender research streams“ [FF21].

Arksey & O'Malley distinguish five stages of the scoping literature review:

- „Stage 1: identifying the research question
- Stage 2: identifying relevant studies
- Stage 3: study selection
- Stage 4: charting the data
- Stage 5: collating, summarizing and reporting the results.“ [AO05]

Regarding stage 1, the research question here is not whether feminist theories are considered in the use of technology in the public sector as in Feeney & Fusi, but how gender is generally understood in e-government research and what kind of empirical evidence there is in the papers examined. To identify relevant studies, the following literature databases were searched on 07 June 2022: Scopus, ACM Digital Library, IEEE Digital Library and SpringerLink. In the search algorithm it was specified that all fields must contain one of the terms "gender women men Geschlecht Frauen Männer" and the term "egovernment". To analyze the results in depth rather than breadth, the time period was limited to all publications in 2021. This clearly reduced the number of search results. It is also important to note that SpringerLink was searched exclusively for articles and conference proceedings, thus excluding other possible literature sources from the outset.

During a first screening, false hits were sorted out. In addition, duplicates were deleted, and posts were removed from further investigation whose full version was not available. Subsequently, the contributions were evaluated according to the criteria listed in Table 1. Some contributions contained the mentioned search terms but had no relation to gender or e-government in terms of content and / or also no empirical results. In addition, there were contributions in which gender was not mentioned. Accordingly, these studies were also screened out.

Inclusion criteria	**Exclusion Criteria**
One of the words in the document is: “gender”, “women”, “men”, “Geschlecht”, “Frauen”, “Männer”	Government Report / Statement etc.
One Word in the Document is “egovernment”	Books
Published in 2021	Paper is not available
Conference or journal paper	
Contribution contains definition, understanding, or empirical result related to gender	

Tab. 1: Inclusion and Exlusion Criteria

In the fourth step, the following information was drawn from the data: author, title, research objective, method, understanding of gender and gender-relevant outcomes. The results of these steps are summarized in the next chapter.

3 Results

Figure 1 shows the individual steps of the literature selection. Initially 272 papers were identified. After the first screening, 36 entries remained. Many of the sorted-out contributions had no relation to e-government, which was already recognizable in the title. The high number of wrong results may be due to the search word "men", as this word is part of many other words like instrument, argument, development, or acknowledgment. The term "gender" also led to incorrect results in German studies because there are some German word sequences that contain "gender." (e.g., „Fragen der“, „überragender“, „fernliegender“). In addition, the term "gender" also appeared in many papers only in the bibliography, but not in the text itself.

Subsequently, the inclusion and exclusion criteria were applied, whereupon the final number of contributions to be examined was reduced to eight. During the application of the criteria, the contributions were already viewed in more detail. It became clear that some of them contained the word "government" but had no reference to e-government. If the paper focused exclusively on smart city, healthcare, or justice, it was also sorted out.

Some of the contributions also had no reference to gender at all. Finally, studies were excluded from the analysis if they mentioned gender once, but the associated understanding was not apparent, nor were there any empirical results segregated by gender.

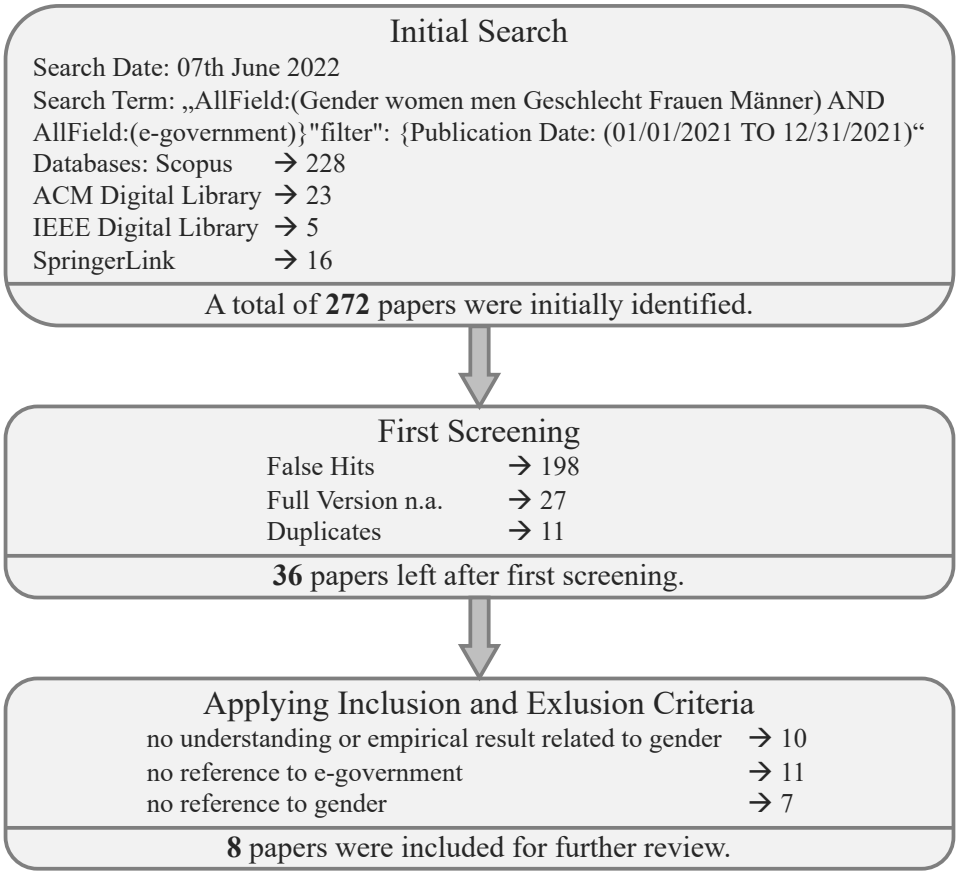


Fig. 1: Overview of the Paper Selection Process

In the next step (Stage 3: study selection, see Chapter 2), the studies were reviewed in detail. Information was then collected on the authors, the title and the research objective, the method, the understanding of gender, and gender-relevant findings. According to Stage 4 (charting the data, see Chapter 2), these information of the remaining eight studies is illustrated in Table 2.

Author & Title	Research Objective	Method	Under-standing of Gender	Gender-relevant Findings
Alarabiat et al.: The suitability of Facebook as a digital platform for constantly perform e-Participation activities	Does Facebook provide all the necessary functions and features to enable citizens to engage in all levels of e-participation?	task technology fit theory, data obtained by survey (N = 211)	binary variable	none
Almuraqab & Jasimuddin: An Empirical Study of Perception of the End-User on the Acceptance of Smart Government Service in the UAE	What factors incluence the end-user acceptance of smart-government services?	partial least squares structural equation modelling (PLS-SEM), data obtained by survey (N = 273)	binary variable	Gender influences the effort expectancy; it has an impact on the relationship between facilitating conditions as well as trust in government and the use of smart government.
Andersson et al.: Citizen Diversity in eGovernment Research	How is diversity addressed in egovernment research?	literature review based on Webster & Watson (2002)	not explicitly stated	From 55 examined papers, 3 focus on gender, two of them suggest no effect of gender on the use of egovernment services; none of the papers included other gender than female and male.
Botrić & Božić: The digital divide and E-government in	What factors of digital divide influence the non-adoption of egovernment in	analysis based on Eurostat Community Statistics on Information	binary variable	Older women are likely to not adopt egovernment services. No effects of gender

Author & Title	Research Objective	Method	Under-standing of Gender	Gender-relevant Findings
European economies	European states?	Society (CSIS) micro data		were found for younger persons.
Janita & Miranda: Exploring critical dimensions for measuring service quality in government to employee	What are the essential characteristics that a university service portal must have to be rated with high quality by employees?	qualitative study (Delphi method); quantitative study (academics from Spanish universities)	binary variable	none
Pietilä et al.: Understanding the youth's user needs for inclusive eParticipation	What are the needs of young people in terms of eparticipation services?	data obtained by workshops: semi-structured group interviews, recording of discussions, and a survey	possibility for participants not to specify their own gender	none
Prokop & Tepe: Talk or type? The effect of digital interfaces on citizens' satisfaction with standardized public service	How does it affect citizen satisfaction when face-to-face interaction is replaced by digital interfaces?	vignette experiment conducted on a sample of German citizens (N = 1.234)	binary variable	Women are slightly more sensitive to technical errors; they are more satisfied with successful service delivery than men; they are also less satisfied when there are service failures without explanation.
Wessel et al.: WCAG and Dyslexia – Improving the	Are requirements of users with dyslexia	development of modifi-cations by literature	third gender available	none

Author & Title	Research Objective	Method	Under-standing of Gender	Gender-relevant Findings
Search Function of Websites for Users With Dyslexia (Without Making It Worse for Everyone Else)	covered by the Web Content Accessibility Guidelines? How can usability be improved for these users?	review and interviews, evaluated by an online survey	called diverse	

Tab. 2: Charting the Selected Papers

In five of the eight posts analyzed, gender was recorded as a binary variable, in one study participants had the option not to specify their own gender, in one post the understanding of gender was not explicitly included, and in one study there was an option for participants to specify "diverse" when asked about their gender. None of the papers address or question current debates about gender; in the majority (five of the eight contributions), the classic binary division is used as a matter of course. In addition, there is another interesting observation: although the variable gender is frequently collected in empirical studies (for example, in surveys or in interviews), this variable is often not considered during further analysis of the results. This is also the case for half of the studies reviewed here. For the four papers where gender-relevant results were found in the analysis, the influence of gender on the research question varied from significant to no influence.

4 Discussion

In accordance with the methodological requirements of a scoping literature review, only a small selection of e-government publications in which gender plays a role could be analyzed in this study. In particular, the restricted time period reduces the total number of studies. Nevertheless, one gets a first impression of the understanding of gender in the respective contributions. For the most part, a binary understanding of gender is assumed. In addition, gender is often asked during data collection but not included in the analysis. However, it has also been shown that there are studies that deviate from this standard. In one study, for example, there was the option of specifying "diverse" as the gender, while in another it was possible not to specify any further information on this. Thus, there is no consistent understanding of gender. On the one hand, this reflects the topicality of the social debate. For in social discourse, too, various positions regarding gender are currently being discussed without any particular view having prevailed. At the same time, a central aspect of scientific work is comparability and traceability, which is made possible, among other things, by consistent definitions. But how can studies be compared if in some there

are two, in others three or more expressions of the variable gender? From this point of view, science has to perform a balancing act in which current social discourses around gender are taken into account, but at the same time the comparability and consistency of definitions is not jeopardized.

5 Outlook

Building on the findings of this work, one could now examine, for example, the extent to which concepts such as *I-methodology* are applied in e-government research and relating technology development. *I-methodology* can be described as the "reliance on personal experience, whereby the designer replaces his professional hat by that of the layman" [Ak95]. In this kind of development process, the artifacts usually demonstrably meet only the requirements of the developers and designers [ORS04]. If this group is not representative of society, the productive systems are only tailor-made for certain people who are similar to the developing persons.

The public sector is also responding to the social discourse around gender. This becomes clear, for example, in new options such as "diverse" or in gender-appropriate language in official documents. It would be interesting to either examine government documents in the form of a scoping literature review or to comparatively analyze artifacts of the public sector with those of academia in terms of their treatment of gender. Finally, how bias is dealt with in technical systems in the public sector and what influence this has on gender could also be investigated.

6 Literature

[Ak95] Akrich, M.: User representations: Practices, methods and sociology. In (Rip, A.; Misa, T.J.; Schot, J.; eds.): Managing technology in society: The approach of constructive technology assessment. Pinter Publishers, London, pp. 167-184, 1995.

[AO05] Arksey, H.; O'Malley, L.: Scoping Studies: Towards a Methodological Framework. International Journal of Social Research Methodology 8 (1), pp. 19-32, 2005.

[De22a] DESTATIS, Statistisches Bundesamt: Studierende in Mathematik, Informatik, Naturwissenschaft (MINT) und Technik-Fächern, https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Tabellen/studierende-mint-faechern.html, accessed: 20/06/2022.

[De22b] DESTATIS, Statistisches Bundesamt: Öffentlicher Dienst. Beschäftigte nach Geschlecht und der Art des Dienst- oder Arbeitsvertragsverhältnisses, Stichtag 30. Juni, https://www.destatis.de/DE/Themen/Staat/Oeffentlicher-Dienst/Tabellen/beschaeftigte-geschlecht.html;jsessionid=044BD46EE2A1018BFF3BBF7D837D882E.live721?nn=212936, accessed: 20/06/2022.

[FF21] Feeney, M. K.; Fusi, F.: A critical analysis of the study of gender and technology in government. Information Polity 26, pp. 115-129, 2021

[Ga21] Gasco-Henandez, M. et al.: Introduction to the special issue on digital government and gender. Information Polity 26, pp. 109-113, 2021.

[LR02] von Lucke, J.; Reinermann, H.: Electronic Government in Deutschland, Speyer Forschungsbericht, 2002.

[MRP01] Mays, N.; Roberts, E.; Popay, J.: Synthesising research evidence. In (Fulop, N.; eds.): Studying the organisation and delivery of health services: Research methods. Routiedge. London, 2001.

[Mu18] Munn, Z. et al.: Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Medical Research Methodology 18, pp. 143-150, 2018.

[ORS04] Oudshoorn, N.; Rommes, E.; Stienstra, M.: Configuring the User as Everybody: Gender and Design Cultures in Information and Communication Technologies. Science, Technology, & Human Values 29 (1), pp. 30-63, 2004.

[St22] Statista: Anteil von Frauen und Männern in verschiedenen Berufsgruppen in Deutschland am 30. Juni 2021, https://de.statista.com/statistik/daten/studie/167555/umfrage/frauenanteil-in-verschiedenen-berufsgruppen-in-deutschland/, accessed: 22/06/2022.

[Tu22] Turner, J.: The 7 Main Ways Technology Impacts Your Daily Life, https://tech.co/vpn/main-ways-technology-impacts-daily-life, accessed: 20/06/2022.

[Un22] United Nations: Achieve gender equality and empower all women and girls, https://sdgs.un.org/goals/goal5, accessed: 20/06/2022.

[VBM10] Viscusi, G.; Batini, C; Mecella, M.: Information Systems for eGovernment. A Quality-of-Service Perspective. Springer, 2010.

Hochschule 2032

Research Data Management in Computer Science - the NFDIxCS Approach

Michael Goedicke[1], Ulrike Lucke[2]

Abstract: This contribution discusses the challenges and architectural considerations for research Data management in computer science and related infrastructure for implementing the so-called FAIR principles. The main challenge is, to cover the research data management requirements of the various sub disciplines of computer science. This diversity must be managed in a uniform way which entails a common structure for this task. We outline these requirements briefly and discuss then the concept of the so-called research data management container (RDMC) which encapsulates a given research data set in conjunction with all accompanying information and support (software, execution environment etc) in order to provide a portable unit for management, distribution and access control.

Keywords: Research Data Management, FAIR Principles, Nationale Forschungsdaten Infrastruktur

1 Introduction

Recently, the discussion of science and scientific processes in general have gained important public attention due to the lack of transparency in certain cases which have been made public. One general factor has been identified in these cases that for a better and transparent scientific process the used data be it measured or generated by e.g. extensive simulation was in many cases not available for scrutiny or even turned out be manipulated. In addition, the desire was apparent to reuse this – in many cases – expensively obtained research data. The actual needs in these areas created baroque effects especially in attributing research data in publications – like including the data owners as authors of papers who had no other part in the actual paper at hand.

These are general observations. In computer science (CS) specific observations regarding research data can be made. If a piece of research in CS generates, observes or uses a data set to support the published results it is was in many cases quite easy to "publish" them: create a specific website for the purpose and put it there. In addition, further artifacts (e.g. software and scripts) and context information (e.g. execution environment) can be included there as well. However, this ad-hoc measures suffer from a range of problematic aspects: the site is often not maintained thus the data and artifacts are not useable after a while and the entire piece of information is not findable in a systematic way due to the

[1] paluno/University of Duisburg-Essen, Faculty of Economics and Business Administration, Gerlingstraße 16, Essen, 45117, michae.goedicke@paluno.uni-due.de,
[2] Complex Multimedia Architectures/University of Potsdam, Department of Computer Science, An der Bahn 2, Potsdam, 14476, ulrike.lucke@uni-potsdam.de

doi:10.18420/inf2022_112

lack of standardized meta-data.

The approach to realize more transparency at a general scientific level created the so-called FAIR[3] (Findable, Accessible, Interoperable, Reusable) principles for research data management. These must be translated into actual activities for each specific scientific discipline. In CS a consortium has been created to define such activities and measure and funding is requested in a proposal (NFDIxCS) as part of the NFDI[4] – initiative. This national initiative has been started to create a research data management (RDM) infrastructure to support scientific communities in a discipline specific way. Thus, we are presenting here the CS-way to build such an infrastructure as has been put forward in a related funding proposal (see also the related website[5] where more detail regarding the consortium can be found).

The overall approach is based on the CS scientific community which is included by reaching out using the existing infrastructure of scientific societies. Here the GI[6] plays a key role. While this is important to build a common understanding of the types of research data and how the much-needed transparency of scientific processes is improved for CS we will concentrate on the infrastructure in this contribution. The accompanying measures to reach out to the CS-community will be also supported by the infrastructure, of course.

Thus, this contribution is an account of the key elements of the proposed infrastructure. We briefly describe how this relates to the FAIR-principles and how basic requirements of our scientific discipline are mapped onto related RDM concepts. Finally, the architecture of the infrastructure is sketched where the so-called Research Data Management Container plays a major role.

2 Transparency in Science: The FAIR Principles for RDM in CS

Based on an intensive discussion in the structures of the GI an understanding the FAIR principles and their application to CS was created. We discuss this implementation of these principles in NFDIxCS briefly below. This will include also measures to quality assurance for research data and services to involve the community continuously which we will not address here. This altogether will enable the cultural change towards FAIR research in CS.

The CS discipline is nowadays very broad since CS also reaches out to quite big range of applications. In this work we address not the applications of CS but the findings and research which relates back to CS from applications e.g. search in biological structures like DNA provides new insights into general searching algorithms. Thus, the entire types of research data become quite diverse and in short covers all kinds of data structures from e.g. unstructured point clouds, semi structured texts like software artifacts or automated

[3] https://force11.org/info/the-fair-data-principles/
[4] https://www.dfg.de/nfdi
[5] https://www.nfdixcs.org
[6] Gesellschaft für Informatik https://www.gi.de

theorem proofs to highly organized relational data in tabular form. We used a categorization of the DFG and the subdivision of CS in the substructures of the GI to create related profiles of research data types in CS. This would fill another paper thus whenever the reader considers her/his own subdiscipline just think of it being covered by the NFDIxCS approach. Below we address this as the *profiles of [CS-] research data*.

2.1 Measures to Implement the FAIR Principles across CS

The FAIR principles are the guide and compass for the work within NFDIxCS. The design of the infrastructure, the services and the processes will always carry the FAIR principles as general requirements. It is clear that these requirements have to be complemented with possible conflicting requirements e.g. in terms of legal, intellectual property rights or privacy requirements, and not all FAIR principles can be realized to the utmost extent. We foresee creating separate views on the metadata and data that satisfy the varying requirements. The use of a specific view is then controlled by the role a user has on the data (e.g. author, reviewer, reader). Also, several rule sets, process models etc. have been set forth to guide RDM in general (see e.g. [RISE] and [DIAMANT]) and support ways to implement the FAIR principles within scientific practice. Here, we address the measures at an abstract level to support and guide the work within NIFDxCS by using these knowledge sources. We walk through the principles to point to the planned measures.

- To be **Findable**: This principle addresses the role and properties of identifiers and metadata which are part of a search infrastructure identifying the data and related context. The general related concepts for CS research data as characterised in the profiles as mentioned above. In order to be able to connect to and interoperate with other national and international partners, specifications need to be standardized between stakeholders beyond NFDIxCS.
- To be **Accessible**: This principle addresses the way research data is actually stored and can be accessed through the metadata and identifiers via a search infrastructure. Related open protocols using open authorization and authentication will be key for this topic. NFDIxCS related work packages will address its realization.
- To be **Interoperable**: Interoperability must be realized at all levels of our research data management architecture (see below). The metadata and related protocols need to be open and shared among the community at large and need to be integrated into the general infrastructure of NFDI and possibly beyond. Thus, the overall architecture, meta data, protocols and common standards address this challenging part of the work. Part of the community building will address the related dissemination.
- To be **Re-Usable**: This principle addresses the way the research data can be reused easily and openly. Plurality of attributes, access license(s), provenance and community standards are the aspects addressed by respective work packages. A special aspect supporting reusability will be addressed in work packages addressing reusable execution environments for the long term and architectures / interfaces. While the

former will provide concepts and implementations of the RDMC concept (see below) the latter work addresses the whole structure where instances of RDMCs can be maintained and provided for storing and accessing the data via the appropriate role-based interfaces.

Also, the composition of the ruling bodies of NFIDxCS as well as the operating model are formed in such a way that the continuous observation of the FAIR principles guides the monitoring, discussion and decision making within NFDIxCS as well. We will design a specific code of conduct for NFDIxCS to explicate our common understanding of the FAIR principles against our disciplinary background. This approach will be based on the existing GI statutes and ethical guidelines, leading to the NFDIxCS bylaws.

2.2 Research Data in Computer Science

The CS community is methodically well-prepared to specify abstract-level descriptions of an entity, here research data and its accompanying information. There are plenty of formal description languages to model, for instance, domain-specific data structures (syntax and semantics), system components (structure and behaviour), environments (input from and output to human and/or technical context) etc.. These vocabularies will be a valuable basis to derive metadata standards by a) categorization of possible dimensions and features of different CS research data and b) relating them to each other within a taxonomy. For these semantic aspects we rely on previous work on the Open Research Knowledge Graph [ORKG]. These quite abstract concepts are more at the level of publications and experiments. However, it can be used to e.g. link papers with the datasets used in the paper. Structures in the ORKG are extensible, which we intend to use for CS-specific approaches to characterize our dedicated types of research data according to the needs of the identified sub-disciplines.

The challenge of metadata management in NFDIxCS is the availability of vocabularies optimally reflecting necessary technical specifications, formats, languages, notations etc. for all categories of data in CS. We compiled a sample of relevant approaches for these categories of data and thus exemplifies the range of items to be covered by metadata vocabularies to be developed for CS.

Table 1 Dimensions for specification of selected CS data (sub)categories

(only samples given in no way a claim for completeness)

Data Category	Data Subcategory	Examples for technical specifications, protocols, formats, languages etc.
Data sets	strongly structured	SQL, RDF, SPARQL, GraphQL, JSON, XML
	semi-structured	CSV, TSV, NetCDF, HDF5, Apache

		Parquet
	unstructured	txt, mp3, mp4, jpg
Software	formal models	UML, BPMN, DMN, ArchiMate, ERM
	scripts	Python, R, MatLab, Visual Basic
	web/microservices	HTTP, REST, JSON, AMQP, MQTT
	complex systems	YAML
Context	operating systems	MS Windows, macOS, iOS, Android, various Unix versions and derivatives
	Programming models & runtimes	C, C++, Fortran, Java, OpenMP, MPI, CUDA
	supporting services	CI, CD, container services, server virtualization
	hardware	Configuration files, specifications, emulators if available/necessary
	Physical location	GPS, Address, location in rack, room no.

It is important to strengthen process-oriented metadata management along the data lifecycle and the re-usability of metadata to reduce the burden of documenting data as well as to assure quality of that metadata. Researchers must be encouraged in their roles as data users and data producers by supporting the documentation of data as early as possible during the preparation of the data collection using custom metadata-tools. An improvement processes will take into account international standards in the field, with a focus on harmonising, internationalising and standardising current developments e.g. in the context of RDM and other related policy-making bodies.

2.3 Summary: The Requirements

Based on the aforementioned profile descriptions of our sub-disciplines, the CS community is dealing with a large variety of data types that is or has to be considered for sharing, archiving and publishing. Besides (1) traditional datasets (usually originating from a specific field of application), our work focuses on (2) software as research data, and the preservation of its (3) environmental conditions (context) as a third category of CS data. Data can be related to each other, e.g. several datasets are linked to the same software for production or processing, or a software requires a certain execution environment. Moreover, data can be associated not only with (4) metadata (as familiar from other disciplines) but also with different forms of (5) documentation (e.g. admin, developer or user manuals

as well as abstract level design artifacts).

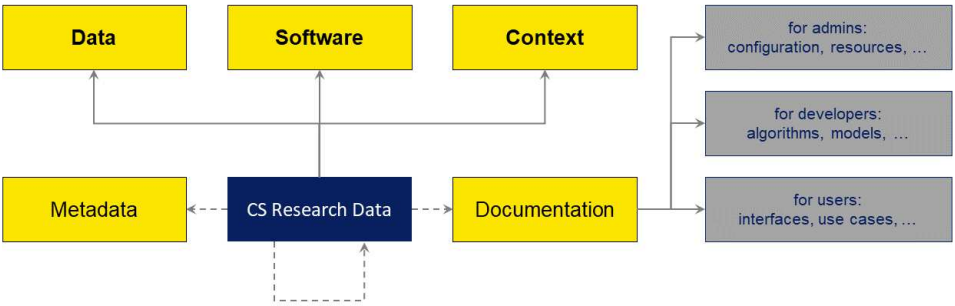


Figure 1 The conceptualization of CS research data in NFDIxCS addresses the characteristics of the discipline in the variety of data types involved.

Independently of the respective sub-discipline, this conceptualization of CS research data (see Fig. 1) can be considered a general model for data handling across CS.

In summary, this leads to the overall architecture of the NFDIxCS infrastructure as presented below derived from the profile descriptions of the sub-disciplines and their declared RDM needs and solutions – for shaping and re-organizing RDM in CS.

3 The Architecture and the RDMC

The major goal of NFDIxCS is to increase sustainability and reusability of research results in CS. Thus, FAIR research data is intended to become an integral element of CS research in general and its publication culture in particular. The core of the efforts is thus to build an infrastructure which addresses the goals of the community as discussed above.

Below we sketch the architecture of the NFIDxCS services as well as the important constituent component: the RDMC for storing and accessing the research data plus accompanying information.

3.1 The Overall Architecture of the NFDIxCS Infrastructure

The infrastructure of NFDIxCS is a stratified set of services. These services are hosted by service providers within the consortium and centrally managed by the NFDIxCS executive management group.

The Access Services (see Fig. 2) provide the direct ability to search, use the search results in form of access to the research data and, in addition, quite a range of (semi-) automated community and management services to NFDIxCS users including central first contact

points for help and support. There is also a machine level interface facilitating the outreach to other organisations within NFDI and beyond, nationally and internationally.

In the middle of the architecture, a set of Core Services provide dedicated functionality to establish the FAIR principles, to support community processes and necessary management operations like ID management and monitoring.

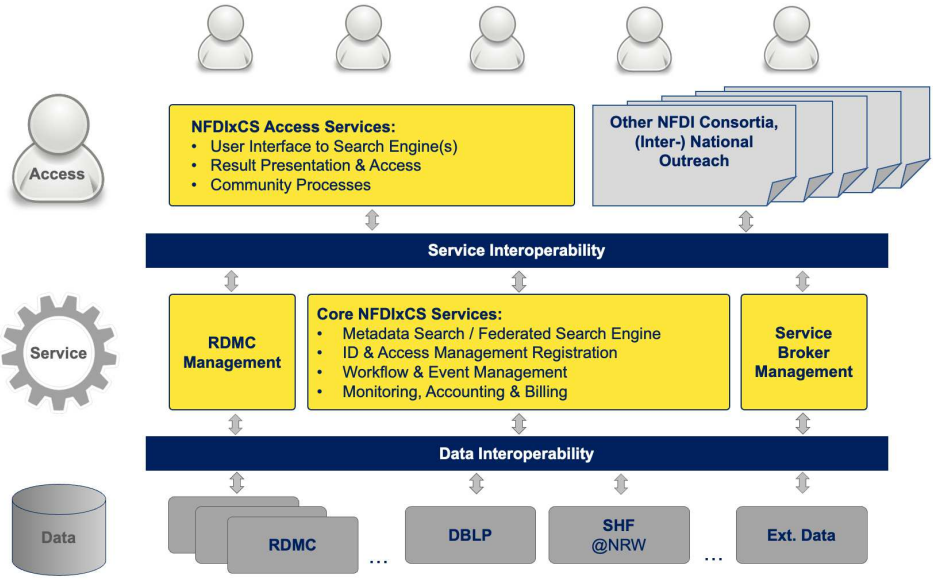


Figure 2 Overview of NFIDxCS Services

At the bottom of the architecture, a set of discipline-specific Data Services are provided to store and access research data and other related databases and repositories. Here, we will include the services of dblp [dblp], facilitating the improved CS related publication processes, and the services of the Software Heritage Foundation [SHF] which are provided as a mirror by the University of Duisburg-Essen to give access to all needed software including all past versions for long term storage (SHF@NRW). Other external repositories and services will be integrated in the NFDIxCS architecture to enable cooperation with other NFDI consortia (using the agreed container interfaces) or basic services like DFN-AAI for identity and access management, for example.

Research data to be included into NFDIxCS will be available in a packaged format: the so-called Research Data Management Container (RDMC), which is a portable object that manages itself in terms of access, specific workflow and all the data, software and further information to describe the data and all the means to access / bring the data back to life. These RDMCs will be hosted by a few service providers initially within the consortium and later, especially after the funding periods, by trusted (probably also paid) service providers which will possibly be available by EOSC / GAIA-X or similar provisioning

schemes. The components encapsulated by an RDMC are sketched in Fig. 3.

We are aware that the idea of encapsulating research data in a container gains much attention in the community (e.g. code ocean, GIDA , Fair Data Objects etc.). We observe a very dynamic situation and currently each of them has a certain profile and is in a different state of maturity. Of course, the RDMC is also the unit for evolution should a specific technology be no longer supported and the entire set of research data and accompanying information needs to be moved to a new set of technologies.

The CS sub-disciplines will provide – managed by the overall compliance and consultancy processes of NFDIxCS – templates for the different research data types in CS. The ability to store the raw and processed data together with (a reference to) the needed software and execution context has been mentioned already ([Str21] and many others). In addition to this, the special NFDIxCS feature are the three components (Access Control, Workflows and Filters & Transformation) on top, which rule the access and the work state of the data, including filters and role specific transformation within the container.

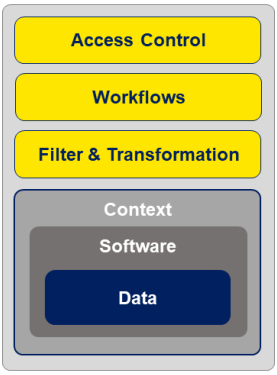


Figure 3 Overall Structure of a Research Data Management Container (RDMC)

Given the envisaged service structure and the RDMC concept, we now go into detail, which aspects of CS RDM we will address in NFDIxCS. Based on such an infrastructure, we support the current research methods and the types of data that typically are generated

3.2 The RDMC

A centrepiece of the storage and management of the actual research data management is the concept and realization of the RDMC. This provides an encapsulation facility to support and execute the rules which will be designed by NFDIxCS to achieve the goal of the project for each specific research data set being eligible for storage and access by the consortium's resources.

It contains everything to encapsulate the research data in question, the software

components (or reference the specific version or variant in a separate repository as, e.g., the Software Heritage repository) and information on the run-time execution environment and further context necessary to support common tasks like the replication of the creation, purging, analysis and visualization of the data.

Essential features of a RDMC include role-based access control to the encapsulated resources. Additionally, there are plug-ins included to provide proper protection of IPR, privacy and security and potential transformation, depending on context, role and purpose of the access in question. If needed, different specific views must be provided to realize possibly divergent requirements. The general structure of a RDMC is sketched in Fig. 4.

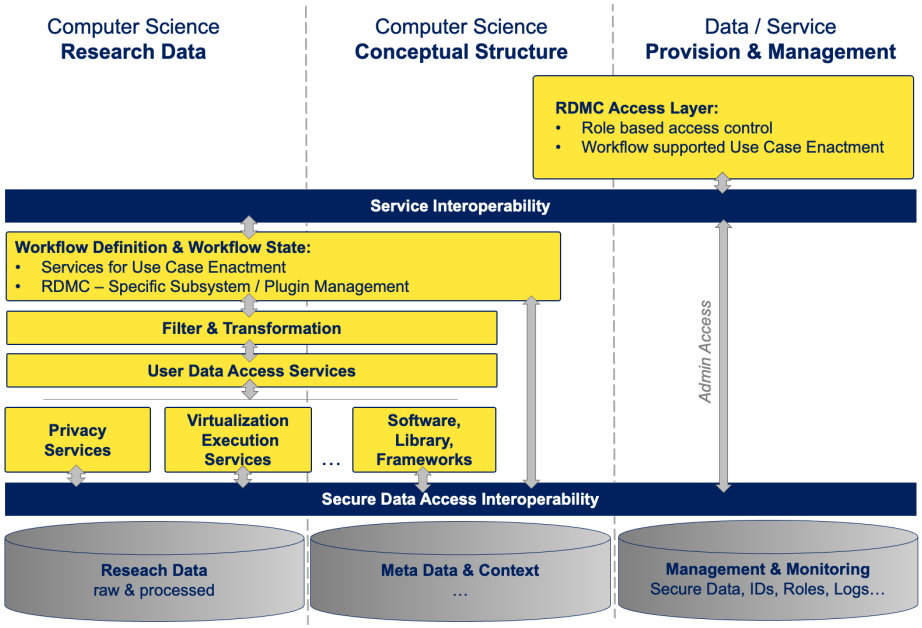


Figure 4 General Structure of Research Data Management Containers (RDMCs) and their management across the levels of the NFDIxCS architecture

For each type of research data identified and described by a CS sub-discipline a type of RDMC is defined and created as a template to collect, store and index a certain dataset with all its belongings as sketched above. Depending on this type, the components as depicted in Fig. 4 might slightly adapt their behaviour to the characteristics of the respective data:

- The RDMC Access Layer restricts the access to the container's resources according to the user's permissions. It uses the same service for access control as the infrastructure.

- The container's workflow defines its possible use cases, e.g. review or reuse processes. The related workflow execution uses the workflow engine from the infrastructure, but the state of the workflow is encapsulated in the container.
- Filter and transformation are plugins which get configured and used based on the use case realized in the container, for instance to implement pseudonymization, anonymization and/or differential privacy. Other transformations, e.g. aggregation of data in case of IPR protection or measures to implement simple privacy protection by eliminating personal data in log files, can be implemented as a plugin here.
- User Data Access Services provides consistent access to the various forms of research data to comply with the given privacy rules defined for the general type of research data and for the specific container in question in addition. Of course, in order to support review in exceptional cases (e.g. a legal inquiry) or finishing a RDMC off at the end of its defined life cycle, admin access is provided as well - secured by additional means like four eyes principle or adequate alternatives.
- Further components help to run the software to access the data using the components privacy services, virtual execution services and additional libraries. This can also include hardware emulation services like Qemu.

Additional considerations are necessary for management of such RDMCs. For instance, it is compulsory for each specific RDMC that the metadata will be made available through the search engines sketched in the overall architecture above. The RDMC Management implements the services to realize the RDMC lifecycle from its inception, registration in the various repositories of the NFDIxCS infrastructure to find it, make it accessible and interoperable and reuse it.

3.3 Sample User Story of the RDMCs

To illustrate the technical considerations above, we provide a sample user story of how the RDMC concept can be used in the realm of the NFDIxCS infrastructure.

In studies on HMI research data is collected in the form of measurements over time. Typically, human participants have to perform a given task for testing the given hypotheses, e.g. on the performance of different interaction techniques. This may be done by capturing data in the system (events like changing a viewpoint, selecting an object or manipulating its characteristics, along with the related timecode) as well as data on the participants (spatial data like movements of the person and his/her hands or head, sensor data like direction of gaze, skin conductance or heart rate). The nature of this data and such studies brings up several aspects to be considered when publishing the data. To name just three:

- Gathered data is closely related to the environment that was used in the study. This includes software aspects, e.g. the used 3D engine or 3D model, the integrated measurement component or the applied analysis tool. Moreover, this in turn includes execution environments, e.g. the used host computer (hardware, operating system,

graphics frameworks and libraries) and virtual reality equipment. For replication of the study to either verify the results or modify the parameters, availability of this software and hardware (at least in an emulated way, as far as this is possible) is a necessary condition. Thus, our RDMC approach encapsulates the data along with these components in order to publish them.

- Involvement of human participants and capture of motion data or physiological data require additional efforts for conservation of privacy. As an example, biometric information could be obtained from typical movement or gaze patterns, making it possible to identify this person. Thus, the data must be anonymized before publication. The required effort needed – pseudonymization, anonymization, additional differential privacy – depends on the actual study and experimental design as well as the scenario of use for this data. Our RDMC approach is equipped with specific plugins and workflow models to implement transformations and filters on the data. Thus, e.g. for quality assurance in the publication process, reviewers might get full access to the raw data. Upon acceptance for publication, data is only accessible in an anonymized way.
- The developed technology or interaction technique might be associated with certain intellectual property rights. Thus, authors might refrain from publishing the whole dataset to the broader public. However, selected parts of the data or an aggregated version might still be worth publishing beyond academic reuse. This depends on the license to be issued for the data and is associated with, again, dedicated transformations of the dataset for publication and reuse.

The figure below depicts these possibilities using the example from HMI sketched above in accordance with the layers of the RDMC as familiar from Fig. 3.

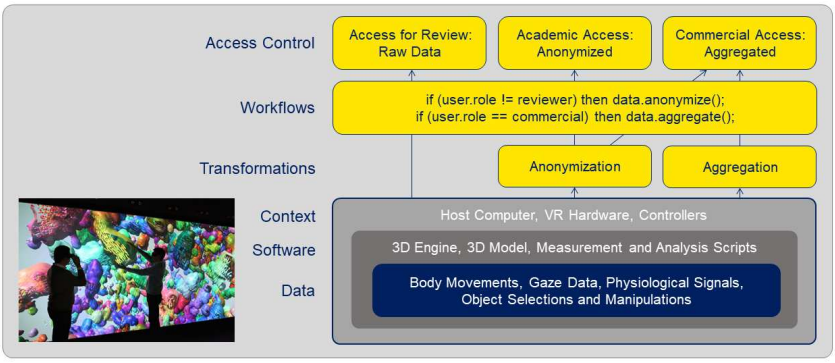


Figure 5 Example of Interaction Data and PlugIns in an RDMC

This way, a trust relation is established between the various stakeholders (authors, reviewers and users of the data) in the research process to ensure that the scientific contribution fulfils the community standards. Authors get their results published in high quality form without harming the privacy of their study participants or their own exploitation interests.

and reviewers can perform a quality check. Users can check and replicate the results to a greater or lesser depth, depending on their role in the academic system – without getting access to the raw data.

Thus, the RDMC is a versatile envelope which provides a great range of configuration possibilities for fulfilling the identified RDM requirements for CS.

4 Summary and the Way ahead

It is obviously quite a work programme for the consortium and the CS community as well. First little experiments have been done to establish proofs of concepts and outreach to national and international societies are developing.

This text is in parts adapted from the proposal for NFDIxCS. Of course, this is a multi-person effort but the responsibility for this text is by the authors.

We are grateful to all who participated and supported us. These are the especially the coapplicants and the participants of NFDIxCS.

We are looking forward to a rewarding journey and to meet you and your RDM requirements.

Bibliography

[DIAMANT] The DIAMANT-Model 2.0 Reference Process https://ubt.opus.hbz-nrw.de/frontdoor/index/index/docId/1432

[dlbp] Computer Science Bibliography https://dblp.org/

[ORKG] Open Research Knowledge Graph https://www.orkg.org/orkg/

[RISE] The RISE-DE Reference Model https://zenodo.org/record/3585556#.X3DX5C-21VQ

[SHF] Software Heritage Foundation https://www.softwareheritage.org/

[Str21] Strickroth, S.; Bußler, D.; Lucke, U., Container-based Dynamic Infrastructure for Education On-Demand. In: Kienle, A., Harrer, A., Haake, J. M. & Lingnau, A. (Hrsg.), DELFI 2021. Bonn: Gesellschaft für Informatik e.V. (p 205-216). https://doi.org/10.1365/s40702-021-00771-7

Kooperativer Aufbau und nachhaltiger Betrieb der Landesinitiative für Forschungsdatenmanagement in Brandenburg (FDM-BB)

Heike Neuroth[1], Janine Straka[2], Carsten Schneemann[3] und Daniela Mertzen[4]

Abstract: Die Landesinitiative Forschungsdatenmanagement in Brandenburg (FDM-BB), ein Verbund der acht forschenden staatlichen Hochschulen Brandenburgs, verfolgt seit 2019 das Ziel, nachhaltiges Forschungsdatenmanagement in Brandenburg zu institutionalisieren. Dazu wurde zu Projektbeginn eine Umfeldanalyse regionaler und (inter-)nationaler Forschungsdaten-Initiativen und Projekte sowie eine Bedarfserhebung zum Forschungsdatenmanagement an den einzelnen Hochschulen durchgeführt. Basierend auf diesen umfassenden Vorarbeiten konkretisierte die Landesinitiative Handlungs- und Implementierungsempfehlungen für das FDM in Brandenburg. Zur Verankerung eines institutionalisierten und nachhaltigen Forschungsdatenmanagements in Brandenburg ist ein Projektantrag gestellt worden, der eine anteilige Förderung durch BMBF und MWFK vorsieht. Innerhalb der Projektlaufzeit sollen, koordiniert durch die Landesinitiative FDM-BB, insbesondere lokale FDM-Kompetenzen an den einzelnen Hochschulen aufgebaut und verstetigt sowie kooperative FDM-Dienstleistungen und IT-Dienste bereitgestellt werden.

Keywords: Forschungsdaten, Forschungsdatenmanagement, Landesinitiative, Brandenburg

1 Einleitung und Kontext

Transparente und nachhaltige Forschung setzt einen systematischen Umgang mit Forschungsdaten, geleitet von den *FAIR Data Principles* [Wi16], voraus. Dem Forschungsdatenmanagement (FDM) wird daher an internationalen sowie nationalen Forschungseinrichtungen eine immer größere Bedeutung zugeschrieben. Auf internationaler Ebene stellt beispielsweise die European Open Science Cloud (EOSC)[5] eine disziplinübergreifende digitale Forschungsinfrastruktur dar, welche Forschungsdaten sowie Dienste und Werkzeuge zum nachhaltigen Datenmanagement zugänglich macht. Auf nationaler Ebene greift die Nationale Forschungsdateninfrastruktur (NFDI)[6] dieses Konzept auf. Deutschlandweit haben sich bereits FDM-Landesinitiativen[7] gebildet, deren

[1] Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Kiepenheuerallee 5, 14469 Potsdam, heike.neuroth@fh-potsdam.de, https://orcid.org/0000-0002-3637-3154
[2] Universität Potsdam, Universitätsbibliothek, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, janine.straka@uni-potsdam.de, https://orcid.org/0000-0002-0695-1689
[3] Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Kiepenheuerallee 5, 14469 Potsdam, carsten.schneemann@fh-potsdam.de, https://orcid.org/0000-0002-2683-5853
[4] Universität Potsdam, Universitätsbibliothek, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, daniela.mertzen@uni-potsdam.de, https://orcid.org/0000-0003-4471-9255
[5] https://eosc-portal.eu/
[6] https://www.nfdi.de/
[7] vgl. z. B. https://www.forschungsdaten.info/fdm-im-deutschsprachigen-raum/deutschland/

doi:10.18420/inf2022_113

Strukturen und Aktivitäten auf Landes- und Hochschul-spezifische FDM-Bedarfe ausgerichtet sind.

Das Flächenland Brandenburg hat mit seinen acht forschenden staatlichen Hochschulen insgesamt relativ wenige Hochschulen: Vier Universitäten[8] und vier Hochschulen für angewandte Wissenschaften (HAW) bzw. Fachhochschulen.[9] Daneben gibt es in Brandenburg eine hohe Anzahl an außeruniversitären Forschungseinrichtungen wie Max Planck Institute, Einrichtungen der Leibniz-Gemeinschaft und der Helmholtz-Gemeinschaft Deutscher Forschungszentren e. V. Alleinstellungsmerkmal der Landesinitiative Forschungsdatenmanagement in Brandenburg (FDM-BB) ist, dass sich diese acht Hochschulen von Beginn an gleichberechtigt engagieren. Die enge Zusammenarbeit der Hochschulen wird sowohl durch regelmäßige Treffen einer Arbeitsgruppe AG-FDM-BB als auch durch eine kollaborative Arbeitsumgebung gefördert. Auf Leitungsebene kooperieren die Hochschulen in einem Steuerungsgremium, in dem alle Vize-Präsident*innen der Hochschulen die Aktivitäten von FDM-BB begleiten und wissenschaftspolitisch unterstützen. Zum erweiterten Netzwerk FDM-BB zählen von Beginn an auch die außeruniversitären Forschungseinrichtungen mit ihrer domänenspezifischen Expertise sowie ihren nationalen und internationalen Netzwerken.

Das Zentrum für Digitale Transformation (ZDT)[10], welches 2019 in Brandenburg gegründet wurde, koordiniert gemeinsame Digitalisierungsprojekte in Forschung, Verwaltung und Lehre und hat das Thema Forschungsdatenmanagement in seinem Bearbeitungsportfolio aufgenommen. Parallel ist FDM-BB mit der Brandenburgischen Landeskonferenz der Hochschulpräsidentinnen und -präsidenten (BLHP)[11] turnusmäßig im Austausch. Die BLHP veröffentlichte im Mai 2021 gemeinsam mit dem Ministerium für Wissenschaft, Forschung und Kultur (MWFK)[12] eine "Gemeinsame Digitalisierungsagenda des Ministeriums für Wissenschaft, Forschung und Kultur des Landes Brandenburg und der brandenburgischen Hochschulen",[13] in der u. a. die Weiterentwicklung von Forschungsdatenmanagement an den Hochschulen festgehalten wurde [vgl. MW21, S. 79].

Die Landesinitiative FDM-BB hatte sich den folgenden Schwerpunkten gewidmet, um ein nachhaltiges Forschungsdatenmanagement kooperativ zu initiieren:

Umfeldanalyse: Analyse der Erfahrungen von FDM-Landesinitiativen sowie Auswertung relevanter nationaler und internationaler Entwicklungen, Initiativen und Projekte.

[8] Universität Potsdam, Filmuniversität Babelsberg Konrad Wolf, Europa-Universität Viadrina und Brandenburgische Technische Universität Cottbus-Senftenberg
[9] Fachhochschule Potsdam, Hochschule für nachhaltige Entwicklung Eberswalde, Technische Hochschule Brandenburg und Technische Hochschule Wildau
[10] https://zdt-brandenburg.de/
[11] https://blhp.de/
[12] https://mwfk.brandenburg.de/mwfk/de/
[13] https://mwfk.brandenburg.de/sixcms/media.php/9/MWFK_Digitalisierungsagenda_21-05-17_BARRIEREFREI.pdf

Bedarfserhebung: Bewusstseinsbildung auf Leitungsebene für die Bedeutung von FDM sowie Identifikation von FDM-Bedarfen an den einzelnen Hochschulen (Arbeitsebene) und systematische Erhebung struktureller Rahmenbedingungen für FDM (Leitungsebene).

Handlungs- und Implementierungsempfehlungen: Gemeinsame Abstimmung und Priorisierung über Handlungsfelder sowie Potenziale möglicher Kooperationsfelder, insbesondere bei wissenschaftlich-technischen FDM-Diensten.

Flankierende Maßnahmen: Verabschiedung einer Forschungsdatenstrategie für Brandenburg, im Bereich Qualifizierung der Kompetenzaufbau für FDM an jeder Hochschule unter Berücksichtigung der lokalen fachwissenschaftlichen Spezifika und finanziellen bzw. personellen Rahmenbedingungen sowie Vernetzung mit (inter-) nationalen FDM-Initiativen und Projekten.

2 Bisherige Ergebnisse und durchgeführte Maßnahmen

2.1 Umfeldanalyse

Die zu Projektbeginn durchgeführte Umfeldanalyse [Wu21] diente der Verortung der entstehenden Forschungsdaten-Landesinitiative FDM-BB auf mehreren Ebenen (vgl. Abb 1). Für die horizontale Verortung als eine von jener Zeit sieben vergleichbaren FDM-Landesinitiativen wurden die bereits bestehenden Angebote, Services und Strukturen analysiert. Über die einzelnen Landinitiativen herausgehend und als vertikale Verortung bezeichnet, wurde der Blick einerseits nach außen, auf FDM-Angebote auf nationaler und internationaler Ebene gelegt, andererseits wurde nach innen, auf die Universitäten, Fachhochschulen und außeruniversitären Forschungseinrichtungen Brandenburgs geschaut, um somit eine möglichst umfassende Umfeldanalyse zu gewährleisten.

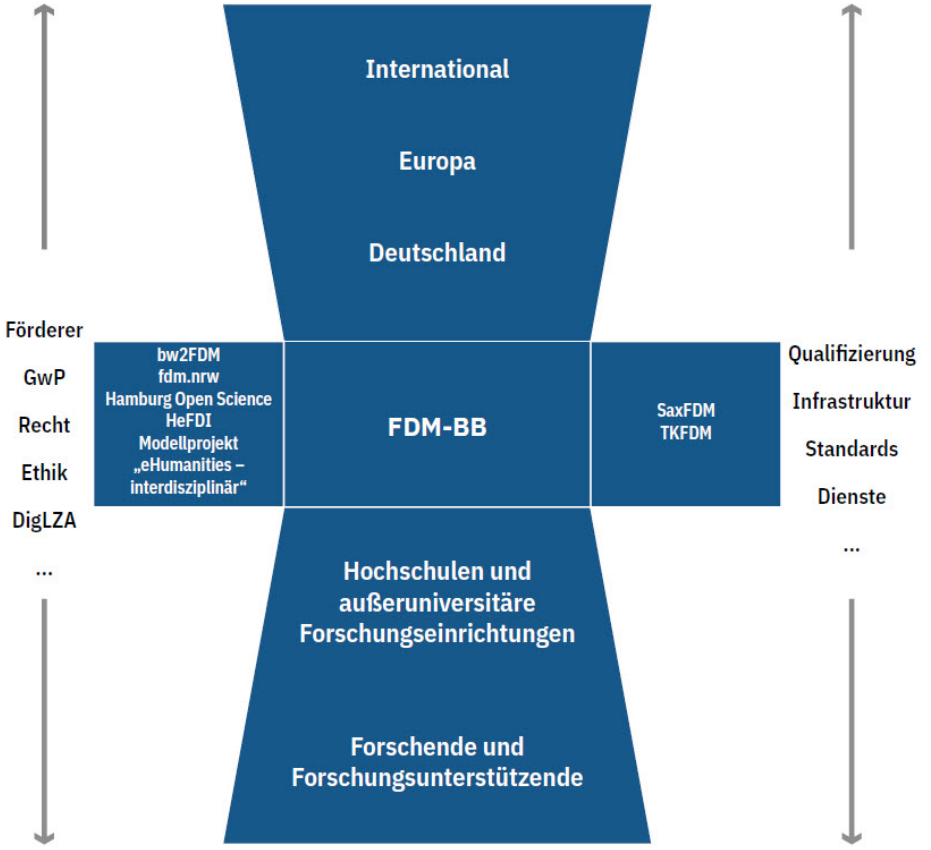


Abb. 1: Vertikal-horizontaler Ansatz der Umfeldanalyse zur inhaltlichen Verortung (Auszug aus Abb. 4, S. 30, [Wu21]).

2.2 Bedarfserhebung

Die Anforderungserhebung der brandenburgischen Hochschulen [Ra20] sollte die Frage beantworten, wie FDM an den einzelnen Hochschulen institutionalisiert werden kann. Relativ früh zeichnete sich ab, dass es in Brandenburg mindestens zwei Geschwindigkeiten bezüglich der FDM-Entwicklung gibt: Die sogenannten „*Early Adopter*“, welche seit einigen Jahren bereits Maßnahmen ergriffen haben und bereits erste Strukturen bezüglich FDM aufweisen sowie eine größere Gruppe, die sogenannten „*FDM-Beginner*“, die erst seit Kurzem, u. a. durch steigende externe Anforderungen, FDM an ihren Institutionen etablieren wollen. Die Erhebung zeigte deutlich, dass es Bedarfe an wissenschaftlich-technischen FDM-Diensten gab, welcher zu diesem Zeitpunkt nicht

abgedeckt werden konnten. Auch im Bereich Personal und Qualifizierung wurden zusätzliche Bedarfe festgestellt, besonders an *FDM-Beginner* Institutionen, an denen FDM häufig ausschließlich als Zusatzaufgabe für einzelne Mitarbeiter*innen ohne FDM-Erfahrung definiert ist. Standortübergreifende Vernetzung der FDM-Verantwortlichen wurde als positives Merkmal hervorgehoben.

Zusammenfassend ergab die Bedarfserhebung folgende Ziele und Wünsche der Hochschulen an eine Landesinitiative: Eine FDM-Institutionalisierung soll begleitet werden von einer Sensibilisierung aller relevanten FDM-Akteur*innen und Forscher*innen. Außerdem befürworteten die Hochschulen Unterstützung bei der Formulierung eigener FDM-*Policies* sowie bei strategischen Entscheidungen. Auch hochschulübergreifende und damit ressourcenschonende Vorgehensweisen im Bereich wissenschaftlich-technischer FDM-Dienste wurden deutlich bevorzugt. Für die Umsetzung dieser Ziele müssen ausreichende Personalkapazitäten sowie Angebote zur Qualifizierung FDM-Verantwortlicher geschaffen werden. Eine stärkere Formalisierung der hochschulinternen Organisation zum FDM bei den *FDM-Beginner* Institutionen wird erst mittelfristig angestrebt. Die meisten Befragten äußerten sich sehr positiv zu kooperativen Lösungen und unterstrichen die Chance für den Aufbau und Betrieb kooperativer Dienste. Von besonderem Vorteil war dabei, dass zu Beginn noch keine Insellösungen bzw. Parallelentwicklungen in Brandenburg etabliert waren. Als wichtig erachtet wurde auch, vorhandene fachbezogene FDM-Kompetenz der außeruniversitären Einrichtungen einzubinden [Ra20].

2.3 Handlungs- und Implementierungsempfehlungen

Auf Basis der oben vorgestellten Umfeldanalyse sowie Bedarfserhebung wurden der Status Quo im Bereich FDM und die Anforderungen der Hochschulen abgeleitet. Aus diesen Analysen sind konkrete Handlungs- und Implementierungsempfehlungen [Ne21] für Brandenburg formuliert worden, die sich an verschiedene Akteur*innengruppen richteten. Eine Priorisierung der nächsten Schritte hin zu einem institutionellen und nachhaltigen FDM konnte mit Hilfe von strukturierten Umfragen an den acht forschenden brandenburgischen Hochschulen vorgenommen werden. Dabei wurde zum einen die Anzahl der Student*innen, Promovend*innen, entfristeter Mitarbeiter*innen nur für FDM, und die Höhe der Drittmittelgelder der vorangegangenen Jahre systematisch erhoben. Zum anderen wurde abgefragt, wo und wie FDM bereits in der Hochschulstruktur verortet ist, welche zentralen Einheiten und weiteren Stabsstellen etc. bereits involviert sind und ob bzw. wo FDM in der Hochschulleitung verortet ist.

Insgesamt konnten durch die Bedarfserhebung (vgl. Kap. 2.2) 28 Themenschwerpunkte identifiziert werden, die sich wie folgt gruppieren lassen (vgl. [Ne21]): (1) Eigenqualifikation und Sensibilisierung hochschulweit, (2) Koordination und Strukturbildung, (3) Information, Beratung, und Schulung, (4) IT-Dienste für FDM und Forschung und (5) Einbindung Verwaltungsprozesse, -strukturen. Die Relevanz für die einzelnen 28 Themenschwerpunkte wurden von den Hochschulen bewertet, so dass sich

ein gemeinsames Ranking über die Priorisierung der nächsten Schritte ergab, welches für die ersten 20 Positionen wie folgt aussieht:

Nr.	Themenbereich	Punkte
1.1	Eigene Information/Qualifizierung/Kompetenzaufbau bei FDM-Verantwortlichen auf Arbeitsebene	31
1.2	Aufbau und Pflege hochschulexternes Netzwerk, Expertenkontakte	28
3.3	Beratung von Forschenden zu Anforderungen der Förderer	25
4.1	DMP-Tool (z. B. RDMO)	24
3.4	Beratung zu und Kommunikation von Best-Practices (z. B. FAIR Data, Metadaten, ...)	23
3.5	Beratung von Forschenden zu technischen Aspekten	23
2.4	Aufbau eines FDM-Kontaktpunkts an der Hochschule, Zusammenführung von vorhandenen Services	22
3.6	Beratung von Forschenden zu rechtlichen & ethischen Aspekten	22
4.2	Archivierungslösung für Forschungsdaten (z. B. RADAR)	22
4.3	Publikation von Forschungsdaten (einschl. DOI-Service)	22
2.2	Erfassung von Bedarfen an der eigenen Hochschule (z. B. über Umfrage oder strukturierte Interviews)	21
2.5	Bindeglied zwischen Forschenden und NFDI-Konsortien bzw. vergleichbaren Netzwerkknoten	21
5.1	Anbindung an FIS/Bibliographie	19
2.6	Identifikation relevanter externer FDM-Angebote, Weiterleitung von Anfragen an externe FDM-Serviceanbieter	18
3.2	Organisation von Informations- und Weiterbildungsveranstaltungen	17
3.1	Erarbeitung und Bereitstellung von hochschulspezifischen Informationsangeboten (Website, Flyer, FAQ, Wiki etc.)	16
1.4	Kontaktaufnahme/Sensibilisierung von anderen Bereichen der zentralen Einrichtungen	14
4.4	Elektron. Laborbuch oder VFU	14
5.2	Finanzierung/Kostenmodell für Dienste	14
1.3	Information/Sensibilisierung von Leitungsebene und Gremien	13

Abb. 2: Ranking der spezifischen Angebote und Dienste (Auszug aus Tab. 1, S. 37, [Ne21]).

Da sich die meisten Hochschulen in Brandenburg noch zu den *FDM-Beginner* Institutionen zählen, prägen diese die Bedarfe stark. Der höchste Bedarf wird (noch) im Bereich des lokalen Kompetenzaufbaus, der Beratung sowie Qualifizierung und Vernetzung gesehen. Aber auch der Bereitstellung von FDM-Diensten wird bereits eine hohe Priorisierung eingeräumt (vgl. Nr. 4.1, 4.2 und 4.3 in Abb. 2). Untermauert wurde diese Priorisierung durch leitfadengestützte Interviews mit anderen geförderten FDM-

Landesinitiativen, um von deren langjährigen Erfahrungen zu profitieren. Die FDM-Verantwortlichen der anderen Landesinitiativen betonten, dass eine intensive und transparente Kommunikation von Beginn an unerlässlich ist, um alle Akteur*innengruppen und Hochschulen in den Prozess einzubinden. Hervorgehoben wurde, dass eine klare strukturelle Verantwortung sowohl auf Arbeitsebene, als auch auf Leitungsebene nicht früh genug vereinbart werden kann. Dies deckt sich auch mit Beobachtungen in Brandenburg und wurde in den letzten Jahren wiederholt im Rahmen von RDA-DE[14] Tagungen geäußert: Oft geht ein Engagement an Hochschulen von einzelnen, engagierten Personen mit zum Teil niedrigen Stellenanteilen für FDM und meist in befristeten Arbeitsverhältnissen aus. So kann ein systematischer Kompetenzaufbau mit z. B. dauerhafter Verantwortung für Beratungsdienstleistungen für Forscher*innen nicht gelingen.

Konkret wurden auf Basis dieser Erkenntnisse Handlungs- und Implementierungsempfehlungen erarbeitet, die sich an den Unterhaltsträger der Hochschulen (MWFK), die einzelnen Hochschulen (Arbeits- und Leitungsebene) sowie an FDM-BB als Vernetzungs- und Koordinierungsinitiative richten. Empfehlungen in Richtung brandenburgisches Ministerium adressierten z. B. klar den Bedarf nach Anschubfinanzierungen bzw. einmaligen Investitionen für den lokalen Kompetenzaufbau an den Hochschulen oder für die Entwicklung und Bereitstellung brandenburgweit abgestimmter IT-Dienste. Empfehlungen an die Hochschulen formulierten den starken Wunsch nach einer definierten strukturellen Verortung von FDM-Zuständigkeiten sowie die Verabschiedung von FDM-Leitlinien, Richtlinien, Handreichungen etc., die die jeweiligen Spezifika der Hochschule berücksichtigen. Der FDM-BB Landesinitiative werden Aufgaben im Bereich Vernetzung, Unterstützung bei gemeinsamen Qualifizierungsangeboten inklusive Schulungsmaßnahmen sowie Bereitstellung von wissenschaftlichen FDM-Diensten wie z. B. den Research Data Management Organiser (RDMO)[15] und das Research Data Repository (RADAR)[16] zugewiesen.

Insgesamt tritt der klare Anspruch zutage, so oft wie möglich gemeinsam im Rahmen der Landesinitiative tätig zu werden, Insellösungen zu vermeiden und kollaborativ wissenschaftlich-technische FDM-Dienste über z. B. Mandantenfähigkeiten zu entwickeln und dauerhaft bereitzustellen. Eine Herausforderung in diesem Kontext dürfte darin bestehen, geeignete Kosten- und Geschäftsmodelle für kollaborative Dienste und Dienstleistungen zu entwickeln und entsprechende Zahlungsmodalitäten zwischen den Hochschulen in Brandenburg zu etablieren.

2.4 Flankierende Maßnahmen

Einige der oben genannten Maßnahmen konnten bereits umgesetzt werden:

[14] https://www.rda-deutschland.de/
[15] https://rdmorganiser.github.io/
[16] https://www.radar-service.eu/radar/de/home

Wissenschaftspolitisch: In der gemeinsamen Digitalisierungsagenda des Ministeriums für Wissenschaft, Forschung und Kultur des Landes Brandenburg und der brandenburgischen Hochschulen [MW21] wurde sich auf die Erarbeitung einer Forschungsdatenstrategie für das Land Brandenburg geeinigt, welche Mitte 2022 veröffentlicht werden soll und auf einem unveröffentlichtem Eckpunktepapier vom Herbst 2021 basiert.

Qualifizierung: Es wurden bereits regelmäßige Schulungen und Workshops angeboten. Beispielsweise wurde die erst kürzlich abgeschlossene 13-teilige Online-Vortragsreihe zum Thema "*Datenschutz in den Stationen des Forschungsdatenlebenszyklus"* für FDM-Verantwortliche angeboten.[17] Zur diesjährigen internationalen "*Love Data Week"* wurden diverse *Coffee Lectures* zu FDM-Themen, insbesondere für Doktorand*innen, Postdocs und FDM-Interessierte, in Kooperation mit weiteren brandenburgischen und Berliner Forschungseinrichtungen angeboten.[18]

Auch der Nachwuchs wird gezielt eingebunden und qualifiziert, so ist beispielsweise im Rahmen des Projektes eine Abschlussarbeit (BA) der studentischen Hilfskräfte entstanden. Sie erstellten einen Stellenkorpus [ZS21b] und die entstandene Abschlussarbeit "*Forschungsdatenmanagement: Arbeitsschwerpunkte für Informationswissenschaftler*innen*" wurde veröffentlicht [ZS21a].

Vernetzung: Die Vernetzung erfolgt über eine Repräsentation bei forschungsdaten.info[19] und forschungsdaten.org[20], RDA-DE, die halbjährlichen Treffen der FDM-Landesinitiativen, GO UNITE![21] sowie den bilateralen Austausch mit Verantwortlichen aus anderen FDM-Landesinitiativen. Dieser bilaterale Austausch hat sich als besonders produktiv erwiesen und mündete z. B. in einem *Memorandum of Understanding* zum Aufbau gemeinsamer Qualifizierungsmaßnahmen mit der Landesinitiative für Forschungsdatenmanagement in Nordrhein-Westfalen (fdm.nrw).[22]

2.5 Aktuelle Situation von Forschungsdatenmanagement an den brandenburgischen Hochschulen

Brandenburg sieht die geringe Anzahl an Hochschulen als Chance, um von Beginn an sowohl auf Arbeitsebene als auch auf Leitungsebene gemeinsam das Thema FDM voranzubringen. Die unterschiedlich großen und teilweise noch jungen Hochschulen weisen ganz eigene Erfahrungen und Kompetenzen bezüglich ihrer Disziplinen und Forschungsschwerpunkte auf. Dieses Wissen wird in der Arbeits- und Steuerungsgruppe zusammen mit den außeruniversitären Einrichtungen gebündelt und somit die unterschiedlichen Ressourcen optimal eingesetzt. Wichtig war und ist die

[17] https://www.dentons.com/de/sebastian-vonhaldenwang
[18] https://www.cms.hu-berlin.de/de/dl/dataman/news/LoveDataWeek22
[19] https://www.forschungsdaten.info/
[20] https://www.forschungsdaten.org/index.php/Hauptseite
[21] https://www.go-fair.org/2021/05/04/go-uni-becomes-go-unite/
[22] https://www.fdm.nrw/

wissenschaftspolitische Unterstützung, in Brandenburg konkret in Form des MWFK, des ZDT und der BLHP. Somit ist FDM in Brandenburg nicht nur durch das Netzwerk fachlich und inhaltlich gut aufgestellt, sondern auch strukturell und politisch fest verankert. Bereits seit Projektbeginn ist das Alleinstellungsmerkmal der Landesinitiative FDM-BB die gleichberechtigte Berücksichtigung aller forschungsaffinen Hochschulen auf Leitungs- und Arbeitsebene.

In einem von der BLHP und dem MWFK getragenen Eckpunktepapier (unveröffentlicht) zur Forschungsdatenstrategie für das Land Brandenburg vom 15.10.2021 heißt es: „Die Maßnahmen und die Weiterführung der bisherigen Anstrengungen der Landesinitiative FDM-BB sollen in weiteren gemeinsamen Drittmittel-Projekten umgesetzt und durch die neuen Hochschulverträge ab 2024 sukzessiv verstetigt werden. Die Hochschulen werden die Verstetigung von FDM langfristig im Rahmen ihrer Globalbudgets absichern."

3 Ausblick: Institutionalisiertes und nachhaltiges Forschungsdatenmanagement

Die bisherigen Ergebnisse und Erkenntnisse (vgl. Kap. 2) bilden eine hervorragende Basis für die weitere Zusammenarbeit. Die Empfehlungen adressieren drei Schwerpunkte, die als Voraussetzung für ein institutionalisiertes und nachhaltiges FDM in Brandenburg identifiziert wurden:

1. Lokaler Kompetenzaufbau an jeder Hochschule, unterschieden zwischen *FDM-Early Adopter* und *FDM-Beginner* Institutionen;
2. Landesweite, kooperative Bereitstellung relevanter FDM-Dienstleistungen und wissenschaftlich-technologischer IT-Dienste;
3. Koordinierung von FDM-BB als Landesinitiative Forschungsdatenmanagement in Brandenburg inklusive Vernetzung, Dissemination und Verstetigung.

FDM-BB begreift FDM als ständigen Prozess, der einerseits durch die Bedarfe der Forscher*innen und Student*innen sowie andererseits durch entsprechende Organisationsstrukturen an den Hochschulen gesteuert werden muss. Landesweite Kooperationen auf allen Ebenen bei maximaler Eigenständigkeit der jeweiligen Hochschule, die Anerkennung der (fach-)spezifischen Bedarfe vor Ort sowie die Berücksichtigung nationaler und internationaler Entwicklungen bestimmen in Zukunft den Weg der Institutionalisierung von Forschungsdatenmanagement an den brandenburgischen Einrichtungen. Die Mehrzahl der brandenburgischen Fachhochschulen bzw. Hochschulen für Angewandten Wissenschaften (HAW) sind stark anwendungsorientiert und interdisziplinär ausgerichtet, so dass z. B. durch die NFDI-Fachkonsortien nicht in allen Fällen passgenaue FDM-Lösungen zu erwarten sind. Daher haben sich alle Hochschulen in Brandenburg in der Landesinitiative FDM-BB auf eine kooperative Arbeitsteilung verständigt, die in den nächsten Jahren aus unterschiedlichen Quellen finanziert werden soll (s. u.). Ausgehend von den unterschiedlichen

Entwicklungsständen und Erfahrungshorizonten an den Hochschulen kann in Brandenburg durch gesteuerte und intensive Nachnutzungen, eng abgestimmten Austausch sowie kooperativ entwickelte IT-Infrastruktur und -Dienste gemeinsam und effizient Forschungsdatenmanagement aufgebaut und nachhaltig betrieben werden.

Die Implementierung eines institutionalisierten und nachhaltigen Forschungsdatenmanagements an allen brandenburgischen Hochschulen bildet den Leitgedanken bei den weiteren Aktivitäten. Jede Hochschule etabliert in den nächsten Jahren geeignete interne Strukturen und Prozesse, um FDM personell, technisch und sachlich[23] zu verankern. Dies berücksichtigt auch interne Organisationsprozesse (*Change-Management*) und Qualifizierungsstrategien, um hauptamtlich tätiges Personal weiterzuentwickeln und adäquat zu qualifizieren. FDM ist z. T. noch nicht überall in einer Abteilung oder zentralen Einrichtung verankert, so dass der Fokus in den nächsten Jahren darin besteht, vorhandene Strukturen zu erproben und Verantwortungsbereiche festzulegen. An anderen Hochschulen müssen die Zuordnungen ggf. evaluiert und justiert werden, abhängig von (neuen) Bedarfsanforderungen der Forscher*innen und Student*innen. Damit dient die Institutionalisierung auch dazu, den jeweiligen Hochschulstandort für zukünftige Forscher*innen-Generationen attraktiv zu gestalten, indem sie ein professionelles Umfeld für FDM vorfinden. Auch die Student*innen in Brandenburg profitieren von der Schaffung bzw. Steigerung von FDM-spezifischem Know-how, indem sie direkt oder indirekt im Bereich FDM und *Data Literacy* gefördert werden.

Bei der Finanzierung für FDM an den Hochschulen für die nächsten Jahre verfolgt Brandenburg verschiedene Wege: So haben sich die Fachhochschulen bzw. HAW in der BMBF-Ausschreibung "zur Förderung von Projekten zum Thema Nachnutzung und Management von Forschungsdaten an Fachhochschulen"[24] mit einem gemeinsamen Verbundantrag beworben. Eine flankierende Projekt-Finanzierung für die Universitäten in Brandenburg wurde seitens des MWFK fest zugesagt, sollte der BMBF-Antrag erfolgreich sein. Ein gemeinsames Arbeitsprogramm mit einer Laufzeit von drei Jahren wurde erarbeitet, welches sowohl anteilig eine BMBF- als auch eine MWFK-Finanzierung vorsieht. Dies unterstreicht den Ansatz in Brandenburg, dass alle Hochschulen eng und abgestimmt im Verbund eine Chance sehen, FDM in der regionalen Fläche zu verankern. Von einer potentiell erfolgreichen BMBF-Beantragung unabhängig, sind zukünftig weitere Anschubfinanzierung für kleinere Vorhaben nötig und sinnvoll. Denkbar für Projektvorhaben sind z. B. Anträge im Rahmen der Digitalisierungsstrategie über das ZDT[25] einzureichen oder gemeinschaftliche kleinere Vorhaben bei der DFG[26] zu beantragen.

[23] Z. B. Lizenzgebühren im Rahmen von RADAR.
[24] https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2021/08/2021-08-17-Bekanntmachung-Fachhochschulen.html?view=renderNewsletterHtml
[25] https://zdt-brandenburg.de/projekte/antrag/
[26] Beispiel Förderprogramm https://www.dfg.de/foerderung/programme/infrastruktur/lis/lis_foerderangebote/forschungsdaten/

Danksagung

Wir danken der Brandenburgischen Landeskonferenz der Hochschulpräsidentinnen und -präsidenten (BLHP) für die fortlaufende Unterstützung sowie dem brandenburgischen Ministerium für Wissenschaft, Forschung und Kultur (MWFK) für die langjährige Unterstützung und Förderung. Desweiteren gilt unser Dank den ehemaligen Projektkolleg*innen, die maßgeblich zum Erfolg von FDM-BB beigetragen haben: Laura Rothfritz, Ulrike Wuttke, Ina Radke, Boris Jacob, Max Kroehling, Miriam Zeunert und ganz besonders Niklas Hartmann, einem der Initiatoren des Projekts bzw. des FDM-BB Netzwerks.

Literaturverzeichnis

[MW21] MWFK: Gemeinsame Digitalisierungsagenda des Ministeriums für Wissenschaft, Forschung und Kultur des Landes Brandenburg und der brandenburgischen Hochschulen 2021, https://mwfk.brandenburg.de/sixcms/media.php/9/MWFK_Digitalisierungsagenda_21-05-17_BARRIEREFREI.pdf, abgerufen 25. April 2022.

[Ne21] Neuroth, Heike; Straka, Janine; Zeunert, Miriam; Schneemann, Carsten; Hartmann, Niklas; Radtke, Ina: Handlungs- und Implementierungsempfehlung zum Forschungsdatenmanagement in Brandenburg (Projektreport FDM-BB, AP 3, Version 1.0). DOI: https://doi.org/10.25932/publishup-50511, abgerufen 25. April 2022.

[Ra20] Radtke, Ina, Hartmann, Niklas, Neuroth, Heike, Rothfritz, Laura, Wuttke, Ulrike, Straka, Janine, Zeunert, Miriam, Schneemann, Carsten (2020): Anforderungserhebung bei den brandenburgischen Hochschulen (Projektreport FDM-BB, AP 2). DOI: https://doi.org/10.25932/publishup-48091, abgerufen 25. April 2022.

[Wi16] Wilkinson, M.D. et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3:160018. DOI: https://doi.org/10.1038/sdata.2016.18, abgerufen 25. April 2022.

[Wu21] Wuttke, Ulrike; Neuroth, Heike; Rothfritz, Laura; Straka, Janine; Zeunert, Miriam; Schneemann, Carsten; Hartmann, Niklas; Radtke, Ina (2021): Umfeldanalyse zum Aufbau einer neuen Datenkultur in Brandenburg (Projektreport FDM-BB, AP 1, Version 1.0). DOI: https://doi.org/10.25932/publishup-48090, abgerufen 25. April 2022.

[ZS21a] Zeunert, M.; Schneemann, C.: Handlungs- und Implementierungsempfehlung zum Forschungsdatenmanagement in Brandenburg. 2021. DOI: https://doi.org/10.25932/publishup-50790, abgerufen 25. April 2022.

[ZS21b] Zeunert, M.; Schneemann, C.: Stellenkorpus – Forschungsdaten(management) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4430688, abgerufen 25. April 2022.

Abkürzungsverzeichnis

BLHP	Brandenburgischen Landeskonferenz der Hochschulpräsidentinnen und -präsidenten (ehemals BLRK)
BLRK	Brandenburgische Landesrektorenkonferenz
BMBF	Bundesministerium für Bildung und Forschung
bw2FDM	Baden-württembergisches Begleit- und Weiterentwicklungsprojekt für Forschungsdatenmanagement
DFG	Deutsche Forschungsgemeinschaft
DigLZA	Digitale Langzeitarchivierung
DMP	Datenmanagementplan
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
FIS	Forschungsinformationssystem
FDM	Forschungsdatenmanagement
FDM-BB	Landesinitiative Forschungsdatenmanagement in Brandenburg
fdm.nrw	Landesinitiative für Forschungsdatenmanagement in NRW
GwP	Gute wissenschaftliche Praxis
HAW	Hochschule für angewandte Wissenschaften
HeFDI	Hessische Forschungsdateninfrastrukturen
MWFK	Ministerium für Wissenschaft, Forschung und Kultur
NFDI	Nationale Forschungsdateninfrastruktur
RADAR	Research Data Repository
RDA-DE	Research Data Alliance Germany
RDMO	Research Data Management Organiser
SaxFDM	Forschungsdatenmanagement in Sachsen
TKFDM	Thüringer Kompetenznetzwerk Forschungsdatenmanagement
VFU	Virtuelle Forschungsumgebung
ZDT	Zentrum für Digitale Transformation in Brandenburg

Including Data Management in Research Culture Increases the Reproducibility of Scientific Results

Christian Riedel [1], Hendrik Geßner [2], Anja Seegebrecht [3,4], Safial Islam Ayon [5], Shafayet Hossen Chowdhury [6], Ralf Engbert [7], and Ulrike Lucke [8]

Abstract: Reproducible research results are among the pillars of sustainable science, and considerable progress has been achieved in this direction recently. However, there is much room for improvement across the research communities. Here we analyze the reproducibility of 108 publications from an interdisciplinary Collaborative Research Center on applied mathematics in various scientific fields. Based on a previous reproducibility study in hydrology, we identify the rate of reproducible scientific results and why reproducibility fails. We identify the main problems that hinder reproducible results and relate them to previous interventions targeting the research culture of reproducible scientific findings. Thus, the success of our measures can be estimated, and specific recommendations for future work can be derived. In our study, the number of publications that allow for at least partly reproducible research results increased over time. However, we see an ongoing need for directives and support in research data management among research communities since issues concerning data accessibility and quality limit the reproducibility of scientific results. We argue that our results are representative of other interdisciplinary research areas.

Keywords: open data, reproducibility, research data management

[1] University of Potsdam, Institute of Computer Science, An der Bahn 2, 14476 Potsdam, Germany, christian.riedel@uni-potsdam.de, https://orcid.org/0000-0001-5154-4153
[2] University of Potsdam, Centre for Information Technology and Media Management, Am Neuen Palais 10, 14469 Potsdam, Germany, hendrik.gessner@uni-potsdam.de, https://orcid.org/0000-0002-7786-2587
[3] University of Potsdam, Institute of Computer Science, An der Bahn 2, 14476 Potsdam, Germany, anja.seegebrecht@uni-potsdam.de, https://orcid.org/0000-0002-9404-8528
[4] University of Freiburg, Institute of Physics, Hermann-Herder-Str. 3, 79104 Freiburg im Breisgau, anja.seegebrecht@physik.uni-freiburg.de, https://orcid.org/0000-0002-9404-8528
[5] University of Potsdam, Institute of Computer Science, An der Bahn 2, 14476 Potsdam, Germany, safial.islam.ayon@uni-potsdam.de, https://orcid.org/0000-0002-0970-0455
[6] University of Potsdam, Institute of Computer Science, An der Bahn 2, 14476 Potsdam, Germany, md.shafayet.hossen.chowdhury@uni-potsdam.de, https://orcid.org/0000-0001-5536-9807
University of Potsdam, Department of Psychology, Karl-Liebknecht-Str. 24-25, 14476 Potsdam, Germany, ralf.engbert@uni-potsdam.de, https://orcid.org/0000-0002-2909-5811
[8] University of Potsdam, Institute of Computer Science, An der Bahn 2, 14476 Potsdam, Germany, ulrike.lucke@uni-potsdam.de, https://orcid.org/0000-0003-4049-8088

doi:10.18420/inf2022_114

1 The need for and measures to increase reproducibility

The reproducibility of research results is one of the cornerstones of scientific quality assurance [De22]. If a published result is reproducible, this lends credibility to the original study. While most research publications in applied mathematics and the sciences are based on research data and software, many scientific communities currently report low reproducibility rates [St19a, Nü18] or mixed results [SSM18, Ob20, CP16]. This concern, sometimes referred to as a *reproducibility crisis*, is well known within the scientific community [Ba16]. However, institutional examinations on reproducibility rates are rarely carried out, particularly regarding research data management (RDM) interventions [Pe17].

The resulting knowledge gap is more remarkable as the measures to increase reproducibility start at the institutional level, including education and training in various stages of the scientific career [WC19] and support and methodological enhancement for dedicated research groups [Gö19]. While such support initially focused on the provision of IT services [WGS16], there is an increasing orientation towards cultural change in RDM, including research interest in this field [Ge21]. This also applies to the Collaborative Research Center (CRC) 1294 – Data Assimilation, which is the subject of this study.

In general, the term *research data* refers to all kinds of data that arise during the planning, implementation, and documentation of scientific projects or are used in such a project [EU19]. This includes measurements, laboratory values, audiovisual information, text corpora, survey data, observation data, methodological test procedures, and questionnaires. In addition, software and simulations - as in the context of our CRC - also represent central results of scientific research and, therefore, are included in this expression. Dedicated skills are required for the systematic handling of research data throughout the whole process of research. Such education and training are also the subjects of current national RDM initiatives[9].

Within this context, the CRC 1294 – Data Assimilation established such support structures. The CRC is funded by the German Research Foundation and involves 15 research projects, where seven projects have a mathematical focus, and eight projects are located in natural sciences. The overall intention of the CRC is to develop advanced numerical data assimilation methods with potential applications arising from other disciplines, mainly within the natural sciences. With the start of the project, the CRC established a framework that supports the handling and exchange of scientific data. This framework includes providing and coordinating IT platforms to facilitate collaborative work, data sharing and archiving, and offering workshops to raise awareness for research data management and the reproducibility of scientific results. The measures go hand in hand with state-of-the-art approaches in research data management in the respective scientific communities.

[9]https://www.mardi4nfdi.de/ and https://nfdixcs.org/ as for our field

By relating the measures to changes in the reproducibility rate, we aim to determine the success of such interventions. To that end, we investigate the reproducibility of 108 scientific results published between 2017 and 2021 by researchers in the CRC. We use this study to investigate why reproducibility fails and give recommendations on institutional measures to increase the reproducibility of scientific results. The structure of this paper is as follows. Section 2 discusses the methodology and results of previous reproducibility studies. On this basis, the study's design and results within the CRC are presented in section 3. We present a detailed analysis of the reasons underlying reproducibility failures and investigate changes in the reproducibility rate over time to relate the observed change to the interventions on research data handling carried out in the CRC. From this correlation, we evaluate the success of these measures. Our study concludes with recommendations (section 4) for further actions transferable to other research areas.

2 Analyzing reproducibility in different fields of science

Investigations on the reproducibility of scientific results have shown that the inability to reproduce scientific findings is a commonly observed phenomenon that affects multiple research fields. Such studies typically attempt to repeat the analyses in peer-reviewed publications from a particular research field, which often involves the analysis of research data. Despite different habits in data handling in the respective communities and slightly diverging definitions of research data and reproducibility, reproducibility studies commonly conclude that the unsuccessful replication of scientific findings is a central problem that needs to be addressed [CL15, CP16, SKL18, Ra19, St19a, Ma20, La22].

Since what scientists perceive as their generated research data varies with different methodological approaches, the assessment of reproducibility criteria typically depends on the research field. In computer science, reproducibility describes the ability of researchers to generate the same results as another research group using their data and experimental setup [ACM20]. In investigations covering various scientific fields, we find slightly different definitions that affect the methodology to evaluate reproducible science. For example, authors may allow for a certain margin when reproducing numerical results [SKL18, La22] or require a data availability statement from a journal publication [St19a]. In the study by [CP16], for example, the authors examined journal and conference publications in computer sciences and defined reproducibility as the ability to build and run computer code without investigating the correctness of the obtained results. Therefore, the authors excluded theoretical papers describing algorithmic approaches that are not backed by computer code. The study by [Ra19], on the other hand, examined the reproducibility of such theoretical papers in machine learning research and translated the presented methods into computer code to investigate barriers in the respective algorithms. A paper was considered reproducible when the

code written by the authors could use reasonable standard programming libraries, and the code confirmed a majority of the claims from the paper.

The studies above show that different perceptions of the term *research data* and different methodological approaches affect the assessment of reproducibility, which inherently introduces subjectivity when considering numerical reproducibility rates. For example, the rates of reproducible scientific results can vary dramatically from less than two [St19a] to more than 60 % [Ra19], depending on the research field considered and methodology approach. However, the reasons why reproducibility fails are reasonably similar. In many cases, it is impossible to access the data and code of other researchers, which significantly decreases the chance of reproducing scientific results [La22, St19a, SSM18, CL15]. If research data is accessible, missing directions on using the given data and code is another substantial issue [Ma20, Ra19, St19a]. Therefore, we consider aiming attention to reasons why reproducibility fails as more beneficial than addressing reproducibility rates alone.

With digital research data playing an essential role in most research fields and technological advances that facilitate access to data sharing platforms, scientific journals and other stakeholders across many scientific disciplines introduced policies on data handling to improve research data sharing. The investigations by [La22, SKL18, SSM18, CL15] show that the reproducibility of scientific results correlates with the availability of research data and that the introduction of policies enforcing obligatory data sharing by scientific journals significantly affects the availability of research data. Further recommendations by the research community to increase the reproducibility of scientific results include

- mandatory code sharing and reproducibility checks by authors or publishers [La22; SKL18; CL15],
- the use of code control systems or individual training to improve code and data readability [Ra19, CP16],
- permanent contact addresses to reach out to researchers familiar with a specific publication [CP16], and
- a change in research culture that rewards code and data quality over citations and publications [St19a].

Our reproducibility study follows the approach by [St19a]. The authors use a questionnaire [St19b] to assess the reproducibility of scientific findings based on data accessibility, data availability, and the correctness of the obtained results. The study allows for systematic identification of reasons for failed reproducibility that agrees with the ACM's definition of reproducible science [ACM20]. Although we extended the questionnaire of the original study for further prospective analysis (see associated dataset), we use the same criteria to investigate the fraction of reproducible scientific publications, assess reasons for failed reproducibility, and compare our findings to the original study.

Following the criteria presented by [St19a], we investigate whether a publication includes a statement on data availability. We do not conduct a reproducibility analysis if the publication does not include a statement where the associated data can be accessed. In order to highlight publication practices in mathematics and related fields, we individually classify theoretical publications describing a new algorithm when they do not include digital artifacts and a data availability statement. Suppose a publication includes a data availability statement. In that case, we attempt to receive data from a publicly available source (e.g., an academic repository) or by contacting the authors or a third party. If the contacted persons did not reply within two weeks after sending a reminder, we did not attempt to investigate the reproducibility further. Once digital artifacts are available, we investigate whether we can reproduce the scientific findings using the given data. If the results only partly reproduce the findings, we consider the results *some reproducible* (see questionnaire in the associated dataset).

3 Results of an analysis over time

3.1 Reproducibility of results

The results of the study are summarized in Fig. 1. The figure shows why the reproducibility of the investigated publications fails according to the criteria *Data accessibility*, *Data availability*, and *Reproducibility*.

Of 108 examined papers, 26 publications are purely mathematical and investigate algorithms for data assimilation without addressing their implementation. Therefore, such publications do not involve the use of digital artifacts. Only half of the 82 publications relying on digital data include a statement on data availability. Accordingly, we merely attempted to receive the research data from 41 out of 108 publications (38 %).

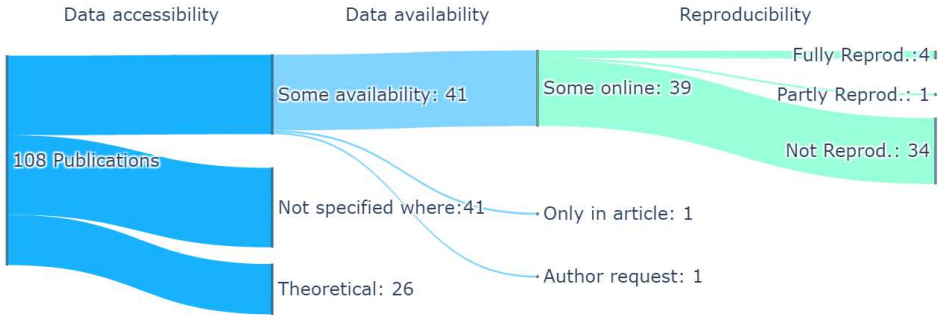


Fig. 1: Sankey diagram showing the results of the reproducibility study based on the criteria Data accessibility, Data availability, and Reproducibility. The down bends indicate at which point we discontinued the study.

We obtained digital artifacts for 39 out of 41 publications (95 %) that include a data availability statement. The data was provided via institutional homepages or data repositories, such as Zenodo, OSF, and Github, or by contacting the authors personally.

Of the 39 publications that have available artifacts, one experiment produces results that correspond to some of the findings presented in the paper but yields an index error during the generation of a plot. Since this error can likely be fixed with some additional code description, we consider this analysis partly reproducible. Overall, five out of 39 publications with data access were fully or partly reproducible (13 %). When considering all 108 publications, 4.6 % of the investigated experiments are at least partly reproducible.

The rate of reproducible research in the CRC is slightly higher than in the original study by [St19a] (see Fig. 2). Analyzing the reproducibility of their investigated publications [St19b] based on *Data accessibility*, *Data availability*, and *Reproducibility* yields that six out of 380 experiments (1.6 %) are at least partly reproducible. When comparing the reasons for discontinuing the investigation in the three respective reproducibility criteria, we see that publications in the CRC have a larger share of theoretical papers and missing data availability statements (62 % vs. 28 %). In contrast, the rate of available artifacts of publications including such statements is much higher in the CRC publications (95 % vs. 60 %). However, the scale of irreproducible scientific results is similarly prevalent despite available artifacts.

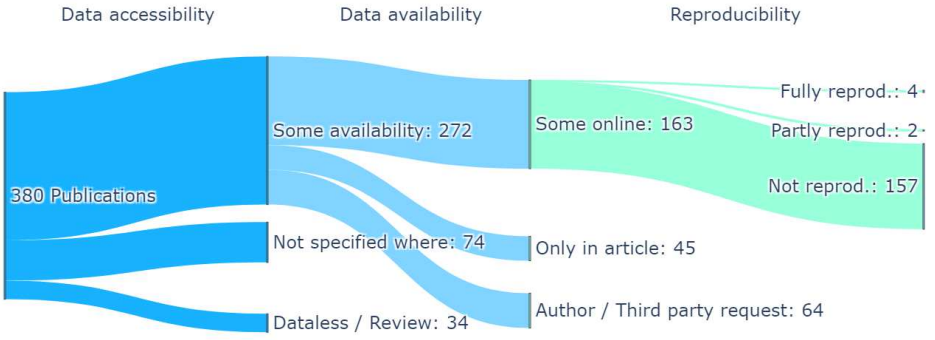


Fig. 2: Sankey diagram showing the results of the reproducibility study by [St19a]

3.2 Reasons for failed reproducibility

Reason for failed reproducibility	Number of affected publications
1. Data Accessibility	**67 (62 %)**
Missing data availability statement	41 (38 %)
Theoretical publication	26 (24 %)
2. Data Availability	**2 (2 %)**
3. Reproducibility	**34 (31 %)**
Incomplete data	20 (19 %)
Software error	8 (7 %)
Insufficient documentation	5 (5 %)
Proprietary software	1 (1 %)

Tab. 1: Reasons for failed reproducibility of 108 papers investigated in this study

Our study results enable the identification of barriers that limit the reproducibility of scientific results. While the original study [St19a] was carried out in hydrology, the CRC publications have a stronger focus on applied mathematics and related application areas. Therefore, the portion of dataless publications that describe advanced numerical approaches is naturally more prominent, leading to a relatively high number of publications labeled as non-reproducible according to our definition. Since translating such numerical techniques into computer code requires extensive mathematical knowledge [Ra19], it is inconvenient for researchers, particularly from other fields, to implement and verify the scientific findings based on the given equations in the respective publications. This circumstance alludes to a data management culture in mathematics and related fields where researchers are often not encouraged to provide digital artifacts for their algorithms.

In order to promote the sharing of research data, several scientific journals have research data policies in place that require authors to include a data availability statement. With the CRC publications, however, a substantial fraction of journals did not expect such a statement at the time of publication. At the same time, many of our researchers did not provide information on data availability on their own accord, leading to a high rate of inaccessible research data.

In contrast to the analysis by [St19a], we were able to receive at least some digital artifacts for most publications that have accessible data. We did not receive the associated data via author request in one case. In another case, the given example code in the publication was insufficient to perform the analysis. However, when artifacts were available, we observed that only in a few cases is it possible to reproduce the published research results. The reasons for failed reproducibility are rather heterogeneous but can, in principle, be classified as *missing data*, *insufficient instructions*, *software errors*, and *inaccessible software*. Most reproducibility attempts in our investigation fail due to missing data. Such failed attempts include incidents where authors did not provide all artifacts necessary to conduct the analysis or where specific resources were not available anymore, for example, because of broken links or lacking details of datasets taken from

data catalogs. Another major limitation was caused by incomplete instructions on how the associated artifacts must be used to reproduce the results. We also encountered several cases where software errors prevented further analysis of the provided artifacts. Such errors refer to weaknesses in the provided computer code, unclear dependencies, or inconsistencies with associated software libraries. In one case, the provided artifacts relied on a proprietary software library that we could not include in our investigation.

Table 1 summarizes the reasons for failed reproducibility based on the criteria *Data accessibility*, *Data availability*, and *Reproducibility*. The *Reproducibility* category includes publications where digital artifacts are at least partly available. Although there are several cases where multiple reasons account for failed reproducibility despite available artifacts (e.g., insufficient documentation and software error), we only characterize the most significant reason that caused our reproducibility investigation to be unsuccessful for each publication. Consequently, we consider the insufficient provision and documentation of digital artifacts and software errors as significant limitations in reproducing scientific results from available data. Furthermore, we attribute deficiencies in data management culture and little motivation to share data as significant limitations to accessing research data.

3.3 Discussion of the results

We set out to investigate and monitor the development of research data management within an interdisciplinary CRC. Our investigation demonstrates that shortcomings in data accessibility and difficulties in reproducing results from available artifacts considerably limit the reproducibility of scientific findings. This issue affects multiple research fields associated with the CRC, and several researchers made equivalent observations [La22, St19a, SKL18, SSM18, CP16]. However, recent recommendations to implement software and reproducibility tests in research [La22, SKL18, CP16] and attempts to push for open data culture in research [St19a, CL15] coincide with the community's demands to promote data accessibility and reproducibility.

Specifically, previous investigations have shown that introducing a mandatory data sharing policy by scientific journals increases research data availability and, therefore, the reproducibility of scientific results [La22, SSM18]. During the period covered by this study, various scientific journals introduced research data policies that make the sharing of associated digital artifacts mandatory. At the same time, there has been a general trend toward open science in many research communities. Accordingly, funding agencies and research institutions implemented corresponding guidelines and recommendations [De22, UP19, ERC17]. Furthermore, since the beginning of its first funding period in 2017, the CRC has provided extended personal support in research data management to scientists. Since 2017, various workshops have been conducted, and support structures have been established to assist and sensitize researchers throughout the research data lifecycle to generate reproducible research results.

We attribute these developments to the positive changes in the CRC's publications that occurred over time, as presented in Figure 3. The figure illustrates the relative annual data accessibility and reproducibility rates for the investigated publications. The data accessibility fluctuates, and we see a moderate decrease in papers that do not include a data availability statement. At the same time, the reproducibility of scientific results increased and four out of 14 investigated papers published in 2021 (29 %) are fully reproducible. Due to a significant increase in reproducible publications with a similar rate of publications sharing data over time, we attribute this progress to several community developments, such as journal policies, but also to the CRC's support structures in research data management established during this period. Therefore, we conclude that implementing advanced data management in research increase the reproducibility of scientific results.

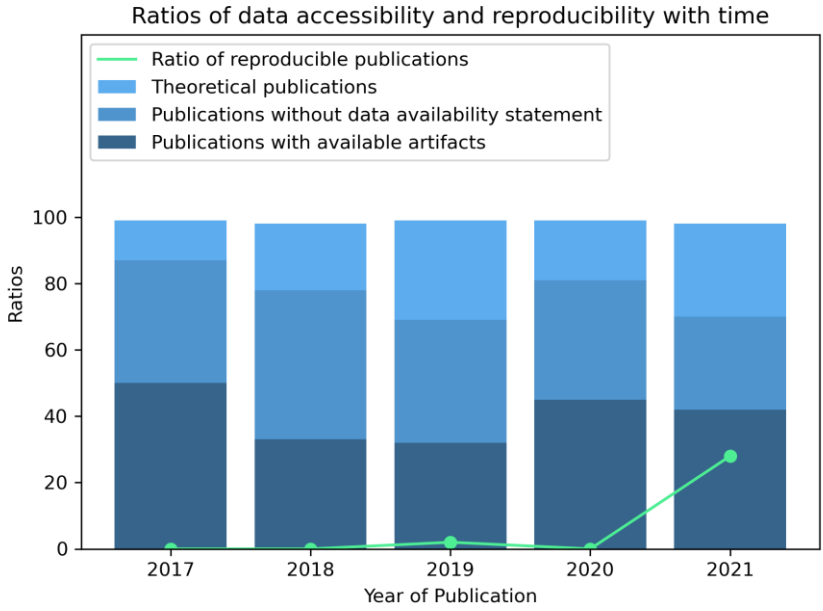


Fig. 3: Annual ratios of the investigated publication’s data accessibility and reproducibility

4 Summary and outlook

Our investigation indicates that significant portions of published scientific findings fail reproducibility checks due to shortcomings in data accessibility and reproducibility of results with available artifacts. Although we have positive experiences with implementing an IT infrastructure to facilitate data sharing and archiving and providing individual assistance in data management to researchers, we see further demand for support structures that promote the accessibility and reproducibility of scientific findings. These structures include support on an individual level, where researchers receive assistance in continuous reproducibility checks and training in data management to ensure adequate data availability and documentation independently.

However, it also addresses research culture in general. For example, we strongly appeal to researchers in mathematics and related fields to support their theoretical publications with computer code. Such data publications facilitate the implementation of numerical techniques and, therefore, increase the impact of numerical methods on other research fields. An interdisciplinary and convenient strategy for this purpose is the introduction of policies that make the sharing of research data mandatory since data availability correlates with reproducibility. Such policies allude to a research culture that recognizes data publications as scientific achievements. We recommend establishing structures that value data quality alongside publications and citations when assessing research quality.

In summary, in order to improve the reproducibility of research results, we recommend to

- establish institutional and nationwide structures that support scientists in data sharing and research data management,
- apply compulsory artifact sharing and reproducibility checks by authors, publishers, and research institutions,
- use version control systems and provide training to researchers to improve the comprehensibility of data and code,
- recognize and reward research data as scientific achievements, including academic hiring processes.

Acknowledgements

This work was funded by Deutsche Forschungsgemeinschaft (DFG) - Project-ID 318763901 - SFB1294. The data supporting the findings of this study are available at https://doi.org/10.5281/zenodo.6543497.

Bibliography

[ACM20] Association for Computing Machinery: Artifact Review and Badging Version 1.1, https://www.acm.org/publications/policies/artifact-review-and-badging-current, accessed: 13/05.2022.

[Ba16] Baker, M.: 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454, 2016. https://doi.org/10.1038/533452a

[CL15] Chang, A. C.; Li, P.: Is Economics Research Replicable? Sixty Published Papers from Thirteen Journals Say "Usually Not". Finance and Economics Discussion Series 2015-083. Board of Governors of the Federal Reserve System, Washington, 2015. http://dx.doi.org/10.17016/FEDS.2015.083

[CP16] Collberg, C.; Proebsting, T.A.: Repeatability in computer systems research. Commun. ACM, 59/3, 62–69, 2016. https://doi.org/10.1145/2812803

[De22] Deutsche Forschungsgemeinschaft: Guidelines for Safeguarding Good Research Practice. Code of Conduct, Version 2, 2022. https://doi.org/10.5281/zenodo.6472827

[ERC17] European Research Council: Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020, https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf, accessed: 13/05/2022.

[EU19] EU Directive 2019/1024 on open data and the re-use of public sector information, 2019. http://data.europa.eu/eli/dir/2019/1024/oj, accessed: 13/05/2022.

[Ge21] Geßner, H.: Transparently Safeguarding Good Research Data Management with the Lean Process Assessment Model. E-Science-Tage - Share Your Research Data, 2021. https://doi.org/10.11588/heidok.00029719

[Gö19] Götzelmann, G.; Hegel, P.; Krewet, M.; Söring, S.; Tonne, D.: Aspects of Digital Infrastructures - Challenges and Perspectives of Research Data in the Collaborative Research Centre "Episteme in Motion". BIBLIOTHEK – Forschung und Praxis, 2019, AR 3284.

[KS13] Kindling, M; Schirmbacher, P.: „Die digitale Forschungswelt" als Gegenstand der Forschung: Lehrstuhl Informationsmanagement. Wissenschaft & Praxis 64(2-3), pp. 127-136, 2013. https://doi.org/10.1515/iwp-2013-0017

[La22] Laurinavichyute, A; Yadav, H.; Vasishth, S.: Share the code, not just the data: A case study of the reproducibility of articles published in the Journal of Memory and Language under the open data policy. PsyArXiv Preprint, 18 Apr 2022. https://doi.org/10.31234/osf.io/hf297

[Ma20] Maassen E.; van Assen, M.A.L.M.; Nuijten M. B; Olsson-Collentine A.; Wicherts J. M.: Reproducibility of individual effect sizes in meta-analyses in psychology. PLOS ONE 15(5), e0233107, 2020. https://doi.org/10.1371/journal.pone.0233107

[Nü18] Nüst, D.; Granell, C.; Hofer, B.; Konkol, M.; Ostermann, FO.; Sileryte, R.; Cerutti, V.: Reproducible research and GIScience: an evaluation using AGILE conference papers. PeerJ 6:e5072, 2018. https://doi.org/10.7717/peerj.5072

[Ob20] Obels, P.; Lakens, D.; Coles, NA.; Gottfried, J.; Green, SA.: Analysis of open data and computational reproducibility in registered reports in psychology. Adv. Methods Pract. Psychol. Sci. 3, 229–237, 2020. https://doi.org/10.1177/2515245920918872

[Pe17] Perrier, L.; Blondal, E.; Ayala, AP.; Dearborn, D.; Kenny, T.; Lightfoot, D. et al.: Research data management in academic institutions: A scoping review. PLoS ONE 12(5): e0178261, 2017. https://doi.org/10.1371/journal.pone.0178261

[Ra19] Raff, E.: A step toward quantifying independently reproducible machine learning research. Proceedings of the 33rd International Conference on Neural Information Processing Systems, Vancouver, pp. 5485-5495, 2019. https://doi.org/10.48550/arXiv.1909.06674

[SKL18] Stockemer, D.; Koehler, S.; Lentz, T.: Data Access, Transparency, and Replication: New Insights from the Political Behavior Literature. PS: Political Science & Politics 51(4), 799-803, 2018. https://doi.org/10.1017/S1049096518000926

[SSM18] Stodden, V.; Seiler, J.; Ma, Z.: An empirical analysis of journal policy effectiveness for computational reproducibility. Proc. Natl. Acad. Sci. 115, 2584–2589, 2018. https://doi.org/10.1073/pnas.1708290115

[St19a] Stagge, J.; Rosenberg, D.; Abdallah, A. et al.: Assessing data availability and research reproducibility in hydrology and water resources. Sci Data 6, 190030, 2019. https://doi.org/10.1038/sdata.2019.30

[St19b] Stagge, J.; Abdallah, A.M.; Rosenberg, D.E.: jstagge/reproduc_hyd: Repository as published in Scientific Data (2.0). Zenodo, 2019. https://doi.org/10.5281/zenodo.2562268

[UP19] Universität Potsdam: Forschungsdaten-Policy der Universität Potsdam, Auszug aus den Amtlichen Bekanntmachungen Nr. 18 vom 30.9.2019, https://www.uni-potsdam.de/fileadmin/projects/ambek/Amtliche_Bekanntmachungen/2019/ambek-2019-18-1361-1362.pdf, accessed 13/05/2022.

[WGS16] Wang, WM.; Göpfert, T.; Stark, R.: Data Management in Collaborative Interdisciplinary Research Projects - Conclusions from the Digitalization of Research in Sustainable Manufacturing. ISPRS Int. J. Geo-Inf. 2016, 5, 41. https://doi.org/10.3390/ijgi5040041

[WC19] Wiljes, C.; Cimiano, P.: Teaching Research Data Management for Students. *Data Science Journal*, 18(1), 38, 2019. http://doi.org/10.5334/dsj-2019-038

Wie steht es um die Sichtbarkeit, Akzeptanz und Nutzung von Datenmanagementplänen an Fachhochschulen?

Qualitative Forschung des Projekts SAN-DMP zur Verbesserung des digitalen Datenmanagements

Mirjam Blümm [1], Christine Burkart[2], Maria Chlastak [2], Katharina Fritsch [1], Heike Neuroth [2], Andreas Schieberle [3], Stefan Schmunk [3] und Stephanie Werner [3]

Abstract: Datenmanagementpläne bilden die Grundlage für einen verantwortungsvollen Umgang mit Forschungsdaten. Gerade (Fach)Hochschulen der Angewandten Wissenschaften stellt dies jedoch vor große Herausforderungen, da sie bislang auf wenig etablierte Strukturen, Standards und Best Practices aus den für sie charakteristischen praxisrelevanten Disziplinen und Kooperationen zurückgreifen können. Das Projekt SAN-DMP will genau in diesem Spannungsfeld ansetzen um die Bedarfe unterschiedlicher Stakeholdergruppen in Bezug auf Datenmanagementpläne (DMP) zu untersuchen. Durch qualitative Methoden sollen evidenzbasierte Umsetzungsszenarien und Handlungsempfehlungen für das Forschungsdatenmanagement entstehen, die auf die substanziellen Bedürfnisse dieses Hochschultyps zugeschnitten sind. Die Expert:innen der verschiedenen Hochschulen und die Forschungscommunity werden schon im Forschungsprozess miteingebunden und die Ergebnisse werden offen publiziert, um die Diskussion und den Wissensaustausch (oder -transfer) anzuregen. Der vorliegende Artikel beschreibt Ausgangslage und methodisches Vorgehen des Projekts.

Keywords: Umgang mit Daten, Datenmanagement, Forschungsdatenmanagement, Datenmanagementplan, DMP, Hochschule für Angewandte Wissenschaften, Fachhochschule, Digitalisierung, Schlüsselqualifikation.

[1] Technische Hochschule Köln, Fakultät für Informations- und Kommunikationswissenschaften & Fakultät für Informatik und Ingenieurwissenschaften, Gustav-Heinemann-Ufer 54, 50968 Köln, mirjam.bluemm@th-koeln.de, https://orcid.org/0000-0003-3665-7031, katharina.fritsch@th-koeln.de, https://orcid.org/0000-0002-0184-7823

[2] Fachhochschule Potsdam, Fachbereich Informationswissenschaften, Kiepenheuerallee 5, 14469 Potsdam, christine.burkart@fh-potsdam.de, maria.chlastak@fh-potsdam.de, https://orcid.org/0000-0003-1046-3634, heike.neuroth@fh-potsdam.de, https://orcid.org/0000-0002-3637-3154

[3] Hochschule Darmstadt, Fachbereich Media, Max-Planck-Str. 2, 64807 Dieburg, andreas.schieberle@h-da.de, https://orcid.org/0000-0001-5389-9409, stefan.schmunk@h-da.de, https://orcid.org/0000-0001-9706-9757, stephanie.werner@h-da.de, https://orcid.org/0000-0002-0468-8856

doi:10.18420/inf2022_115

1 Was sind Datenmanagementpläne?

Es gibt zahlreiche Definitionen, was Datenmanagementpläne (DMP) sind, wozu sie dienen und was andere Personen mit ihnen bewirken (können). Sehr prägnant und anwendungsorientiert formuliert es die Universität Wien:

> *„Datenmanagementpläne (DMP) sind Leitlinien, die den Forscher*innen im Rahmen eines Projekts helfen sollen, adäquat mit Daten umzugehen und eine Nachnutzung zu ermöglichen, darüber hinaus beugen sie auch einem möglichen Datenverlust vor. Durch gezielte Fragen werden Überlegungen zu Verantwortlichkeiten, zum Umgang mit finanziellen und personellen Ressourcen, zu rechtlichen Fragen, Fragen zu Metadaten und Formaten usw. angeregt und mehr Klarheit im Umgang mit Daten gewonnen. Ziel ist, dass Daten langfristig auffindbar, verständlich und nachnutzbar gemacht werden*." [Un22]

Auch die Deutsche Forschungsgemeinschaft (DFG) formuliert Anforderungen bezüglich Forschungsdaten, u. a. zur Beschreibung ihres Entstehungskontextes, zur Einhaltung der FAIR-Prinzipien [FO14] und zu ihrer Ablage in vertrauenswürdigen Repositorien in den Leitlinien 7 und 13 zur Sicherung guter wissenschaftlicher Praxis [Le19]. Dem Auffinden qualitätsgeprüfter Daten und der rechtssicheren sowie einfachen Nachnutzung von Daten kommt in allen Fachdisziplinen und Institutionen eine immer größer werdende Bedeutung zu. Die FAIR-Prinzipien [FO14] [Wi16] bilden dabei disziplinen- und länderübergreifend die Basis für ein gemeinsames Verständnis über den Umgang mit Forschungsdaten nach dem Motto „As open as possible, as closed as necessary" [La20]. Die FAIR-Prinzipien beziehen alle digitalen Daten, die im Zuge von Forschungsprozessen entstehen, ein und berücksichtigen dabei auch Metadaten, Algorithmen, Werkzeuge und Software.

2 Forschungsdatenmanagement

Obwohl DMPs Schlüsseldokumente für das Forschungsdatenmanagement (FDM) darstellen, werden sie bislang nur in Einzelfällen als eigenständige Publikation veröffentlicht, auch wenn eine Vielzahl von Förderinstitutionen DMPs bereits bei der Antragstellung einfordern.[4] Dies stellt zunehmend gerade auch Fachhochschulen (FHs) und Hochschulen für angewandte Wissenschaften (HAWs) vor große Herausforderungen. Insgesamt ist das Bewusstsein für FDM an vielen FHs/HAWs noch nicht besonders ausgeprägt[5], es mangelt häufig an Strukturen wie Beratungsstellen und Infrastrukturen, aber auch an Best-Practice-Beispielen.[6] Fachdisziplinen, die an Universitäten nicht oder weniger praxisrelevant vertreten sind (z. B. Pflegewissenschaft, Rettungsingenieurwesen, Frühkindliche Bildung,

[4] Vgl. z. B. https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2021/09/2021-09-06-Bekanntmachung-Datenkompetenzen.html oder siehe auch [Le19].
[5] Z.B. haben sich an einer bundesweiten Umfrage der Landesinitiative fdm.nrw zu Nutzung von DMPTemplates und -Software (https://www.fdm.nrw/index.php/service/dmp-umfrage/) nur zwei FHs/HAWs beteiligt.
[6] Zu Kriterien für die Evaluation von DMPs vgl. [Pr21].

Design usw.) können zudem nicht auf eine aktive Community zurückgreifen, die fachspezifische Standards entwickelt und Empfehlungen artikuliert. Hinzu kommt, dass FHs/HAWs eine Vielzahl an Projekten zusammen mit Praxispartnern durchführen, zum Beispiel mit Partner:innen aus der (regionalen) Wirtschaft oder Verwaltung, die eigene Bedarfe und durchaus z. B. rechtliche, wettbewerbsspezifische, administrative sowie ökonomische Vorbehalte bezüglich der Veröffentlichung von Daten aus internen oftmals wettbewerbsorientierten Prozessen haben.

Nötig ist größere Transparenz in Bezug auf DMPs, etwa durch Best-Practice-Beispiele, die konkret die Bedürfnisse der FHs/HAWs adressieren. Hierdurch können bei neuen Forschungsvorhaben gerade auch im Verbund mit typischen FH-/HAW-Praxispartnern wie Wirtschaftsunternehmen, Kultureinrichtungen, Verwaltungen, spezifische Orientierungen und Empfehlungen gegeben werden, die insbesondere die Spezifika anwendungsorientierter Forschungs- und Entwicklungsvorhaben im Blick behalten. Dadurch wird es für Forschende und deren Partner:innen auch attraktiver, sich mit FDM auseinanderzusetzen und in den Mehraufwand für DMPs in ihren Projekten zu investieren.

Ziel des vom Bundesministerium für Bildung und Forschung (BMBF) geförderten Projekts SAN-DMP ist es, die Bedarfe und Herausforderungen in Bezug auf DMPs der unterschiedlichen Stakeholdergruppen von FHs/HAWs in einem methodisch reflektierten Evaluationsprozess zu erforschen, um daraus evidenzbasierte Handlungsempfehlungen sowie konkrete Umsetzungsszenarien zur Implementierung von DMPs als zentrales Instrument zur Unterstützung von FDM zu erarbeiten.

3 Methodisches Vorgehen

Um die Besonderheiten von FDM an FHs/HAWs, adressieren zu können, erforscht das Vorhaben SAN-DMP systematisch Anforderungen und Bedarfe der FHs/HAWs an Datenmanagementpläne exemplarisch anhand des Spektrums von drei Hochschulen: FH Potsdam (FHP), Hochschule Darmstadt (HDA) und TH Köln (THK). Die beteiligten Hochschulen decken durch ihre unterschiedlichen Schwerpunkte sowohl in der Kooperation mit spezifischen Praxispartnern als auch im Bereich Forschung & Lehre sowie bei den Akteuren der Forschungsunterstützung eine hohe Bandbreite an Erfahrungskontexten und Praxisbezügen ab.[7] Dazu zählt auch die jahrelange Beteiligung an unterschiedlichen FDM-Landesinitiativen[8]. Um Gemeinsamkeiten, Unterschiede, Hemmnisse und Spezifika

[7] Vgl. dazu auch die an die Hochschulrektorenkonferenz (HRK) gemeldeten Forschungsschwerpunkte der beteiligten Hochschulen, siehe [HR22]. Diese sind: FHP: Digitale Transformation – Urbane Zukunft und Gesellschaft bilden – Frühkindliche Bildungsforschung; HDA: Intelligente Technologien für die digitale Wirtschaft und Gesellschaft – Mobilität und Energie und Verfahrenstechnologien – Nachhaltigkeit; THK: Digitale Arbeit und Lebenswelten – Integrative Gesellschaft und Resilienz – Nachhaltiges Wirtschaften und Ressourcen.

[8] TH Köln mit der Landesinitiative für Forschungsdatenmanagement NRW (fdm.nrw); FH Potsdam mit dem Forschungsdatenmanagement Brandenburg (FDM-BB) und Hochschule Darmstadt mit der Hessischen Forschungsdateninfrastruktur (HeFDI).

in Bezug auf Anforderungen an DMPs analysieren zu können, werden Expert:innengruppen befragt. Die qualitativen, leitfaden-gestützten Interviews dauern rund 90 Minuten und werden via Videokonferenz mit jeweils drei Expert:innen pro Stakeholdergruppe geführt, wobei aus jeder Hochschule eine Vertreter:in je Stakeholdergruppe teilnimmt. Die Erstellung der Leitfragen folgen dem Prinzip „So offen wie möglich, so strukturierend wie nötig“ [He19: 676] und basieren auf dem Konzept der semi-strukturierten Expert:inneninterviews nach Gläser und Laudel [GL19]. Durch die Interviewanalyse werden globale Themen extrahiert [At01], die für eine Szenario-Bildung genutzt werden. Der Analyseprozess bezieht wesentliche diskursanalytische Elemente ein, um den vollen Zugang zu allen relevanten Informationen aus den Gruppeninterviews zu erschließen (vgl. [Da17]). Dazu werden neun Expert:innengruppen-Interviews mit Vertreter:innen aus den drei beteiligten Hochschulen durchgeführt. Diese neun Gruppen stehen exemplarisch für unterschiedliche FDM-relevante Stakeholdergruppen an der Hochschule und repräsentieren die folgenden drei Themen- bzw. Verantwortungsbereiche:

- **Praxispartner**: Wirtschaft, Medien und Verwaltung bzw. Kultur
- **Forschung & Lehre**: (Inter-)Disziplinäre Forschungsprojekte, (Inter-) Disziplinär Lehrende und Studierende
- **Forschungsunterstützung**: Bibliothek, Rechenzentrum/IT-Abteilung und Forschungsservice und -beratung

Das Sampling hat zum Ziel, eine charakteristische Stichprobe zu erstellen, die sowohl regionale als auch bundesweite Besonderheiten beinhaltet und zugleich einen Überblick über die Heterogenität an Praxispartnern von FHs/HAWs darstellt.

4 Erwartete Ergebnisse

Das verwendete Forschungsdesign ermöglicht folgendes: (1) Herausarbeitung der Bedarfe der verschiedenen Stakeholdergruppen an DMPs und Identifikation von Anforderungen, Möglichkeiten sowie Spezifika, (2) Identifikation von Kernelementen eines DMPs, die als Minimalkonsens über Branchen- und Spartengrenzen hinweg definiert werden können, (3) Evaluation von ergänzenden kontextualisierenden Strukturelementen eines DMPs aus Sicht spezifischer Stakeholdergruppen. Der analytische Vergleich innerhalb und zwischen den jeweiligen Stakeholdergruppen der drei Hochschulen, bildet die Grundlage für eine methodisch-reflektierte Kartierung von Lücken, Hürden und Bedarfen an DMPs. Die erarbeiteten Empfehlungen und Umsetzungsszenarien lassen sich für andere FHs/HAWs adaptieren, dadurch einen Wissensaustausch initiieren und an aktuelle Diskussionen und Entwicklungen, wie etwa der Nationalen Forschungsdateninfrastruktur, anschließen.

Literaturverzeichnis

[At01] Attride-Stirling, J.: Thematic networks: an analytic tool for qualitative research. Qualitative Research 2001, I(3), doi: 10.1177/14687941010010030, pp. 385-405, 2001.

[Da17] Davis, C. S.: Focus groups: Applying communication theory through design, facilitation, and analysis. Routledge, New York, 2017.

[FO14] FORCE11: The FAIR Data Principles, https://force11.org/info/the-fair-data-principles/, abgerufen 25. April 2022.

[GL10] Gläser, J.; Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen, 4. Aufl., VS Verlag, Wiesbaden, 2010.

[He19] Helfferich, C.: Leitfaden- und Experteninterviews. In (Baur, N.; Blasius, J., Hrsg.): Handbuch Methoden der empirischen Sozialforschung. Springer Fachmedien, Wiesbaden, doi: 10.1007/978-3-658-21308-4_44, pp. 669–686, 2019.

[HR22] HRK: Forschungslandkarte, https://www.hrk.de/themen/forschung/forschungslandkarte/, abgerufen 25. April 2022.

[La20] Landi, A.; Thompson, M.; Giannuzzi, V.; Bonifazi, F.; Labastida, I.; Olavo Bonino da Silva Santos, L.; Roos, M.: The "A" of FAIR - As Open as Possible, as Closed as Necessary. Data Intelligence 2020, 2 (1-2), doi: 10.1162/dint_a_00027, pp. 47-55, 2020.

[Le19] Leitlinien zur Sicherung guter wissenschaftlicher Praxis. Kodex. DFG, Deutsche Forschungsgemeinschaft, 2019. doi: 10.5281/ZENODO.3923602.

[Pr21] Practical Guide to the International Alignment of Research Data Management—Extended Edition. Science Europe, 2021. doi: 10.5281/zenodo.4915862.

[Un22] Universität Wien. Datenmanagementpläne, https://datamanagement.univie.ac.at/_forschungsdatenmanagement/datenmanagementplaene/, abgerufen 25. April 2022.

[Wi16] Wilkinson, M.D. et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3:160018, doi: 10.1038/sdata.2016.18, 2016.

Erfahrungsbericht aus dem Forschungsdatenmanagement in Bezug auf elektronische Laborbücher

Bert Zulauf [1] und Nina Knipprath [2]

Abstract: An deutschen Universitäten und Fachhochschulen erschließt sich aktuell im Bereich der Struktur und des Aufbaus von Laborarbeit eine neue Arbeitsweise: Dokumentation mittels elektronischer Laborbücher (ELB). Forschende produzieren und nutzen eine bunte Mischung an Daten wie Grafikdateien, Formeln, Tabellen oder Mikroskopie-Dateien. An der Heinrich-Heine-Universität Düsseldorf (HHU) findet regelmäßig Beratung und Unterstützung von Forschenden im Bereich des Forschungsdatenmanagements (FDM) statt. Neben Hilfestellungen bei der Antragsstellung bieten wir auch eine Unterstützung mittels diverser FDM-Tools an, unter anderem auch mit elektronischen Laborbüchern. Warum es sinnvoll ist, elektronische Laborbücher einzusetzen, wie unser Weg damit bisher verlaufen ist und was aktuell zu überwindende Hürden sind, stellen wir in diesem Bericht dar.

Keywords: ELN, ELB, elektronische Laborbücher, Digitalisierung, Forschungsdatenmanagement, Naturwissenschaften, Forschungspraxis.

1 Einführung von ELBs an der HHU

Das Zentrum für Informations- und Medientechnologie (ZIM) der HHU stellte Forschenden bereits seit einigen Jahren das lizenzbasierte ELB "Labfolder" in einer eigenen Hosting Variante zur Verfügung. Im Zuge des vom deutschen Bundesministerium für Bildung und Forschung (BMBF) geförderten Kooperationsprojekts "FoDaKo"[Kn20], einer Zusammenarbeit von drei nordrhein-westfälischen Universitäten, bauen wir nach Evaluation von verschiedenen Softwarelösungen seit 2018 am ZIM mit "eLabFTW" eine Open Source Lösung auf.

Aus infrastruktureller Sicht haben wir seitdem maßgeblich eLabFTW gefördert, da es besser zu den Anforderungen der Forschenden passt. Denn Labfolder geht mit Lizenzkosten einher und ist damit nur schwer für kurzfristige Forschungsprojekte mit Gastwissenschaftlern sinnvoll einsetzbar, ganz zu schweigen von einem kostengünstigen Einsatz in der Lehre, den wir aber ebenso vorantreiben wollten.

[1] Heinrich-Heine-Universität-Düsseldorf, Zentrum für Informations- und Medientechnologie, Universitätsstraße 1, 40225 Düsseldorf, zulauf@hhu.de, https://orcid.org/0000-0002-5747-7892
[2] Heinrich-Heine-Universität-Düsseldorf, Zentrum für Informations- und Medientechnologie, Universitätsstraße 1, 40225 Düsseldorf, knipprath@hhu.de, https://orcid.org/0000-0001-6568-1167

doi:10.18420/inf2022_116

1.1 Technisch-organisatorische Überlegungen

Beide Entscheidungen für die Softwarelösungen fielen aufgrund von Wünschen der Forschenden speziell nach diesen ELBs. Ein wichtiges Kriterium aus Infrastruktursicht war es, vorerst zu einer generischen Lösung zu kommen, mit der eine möglichst große Zielgruppe Forschender gut arbeiten kann. Ein weiterer wichtiger Punkt war, dass es keiner zusätzlichen Software bedarf, um ein ELB zu benutzen, sondern der Service komplett webbasiert angeboten werden kann. Nach einer gleichzeitigen Pilotphase beider Laborbücher entschied man sich dazu, den Fokus auf die Open Source Lösung zu setzen, und nur noch eLabFTW zu betreiben, da diese mehr Anklang unter den Forschenden fand. Deswegen stellte die HHU die Hosting Variante von Labfolder Anfang 2022 ein.

Die Laborarbeit sollte gut strukturiert dokumentiert werden können. Gleichzeitig wollen Forschende auch, vor allem aber in der aktiven Praxisphase, schnell Informationen in einem unstrukturierten Bereich ablegen können, um sie in der Nacharbeit sauber zu dokumentieren. Der ELN-Wegweiser der ZB-MED ist hier eine sehr gute Leitlinie, um sich für eine Softwarelösung zu entscheiden [AL19].

So stellt also die Nutzerseite verschiedene Arten von Anforderungen an ein ELB. Neben den technischen Anforderungen wie Benutzerfreundlichkeit, der Verwendung bekannter Dateiformate und der ständigen digitalen Verfügbarkeit der Daten existiert auch der Wunsch nach Möglichkeit seine Daten komplett wieder aus der ELB Software zu entnehmen und anderweitig weiterzuarbeiten. Hinzu kommen spezifische fachliche Anforderungen, welche sich unter Umständen stark unterscheiden können. Eine der regulatorischen Anforderungen, gerade bei Drittmittelprojekten, ist die Einhaltung der guten wissenschaftlichen Praxis und guten Laborpraxis. Dazu gehört ebenso die sichere Aufbewahrung der Forschungsdaten, wie auch das Sicherstellen der Verfügbarkeit und Auffindbarkeit dieser. Im Bereich der vorklinischen Studien beispielsweise wird eLabFTW als zentrale Schnittstelle für die Komponenten des Qualitätsmanagementsystems und die digitale Speicherung von Laborjournalen verwendet. Es wurden Standardarbeitsanweisungen (SOPs) sowie eine verwaltete Biobank für die sichere Langzeitlagerung von Bioproben eingeführt [He20].

Unveränderlichkeit wird innerhalb eLabFTW durch eine Zeitstempel-Möglichkeit gewährleistet. Hier können Forschende ihre Arbeit zu diesem Augenblick beweissicher festhalten. Eine digitale Möglichkeit Aufzeichnungen über gentechnische Arbeiten nach Gentechnik-Aufzeichnungsverordnung (das sogenannte "Formblatt-Z") ebenfalls elektronisch zu speichern, wird momentan an unserer Universität diskutiert und entwickelt. Forschende in Genlaboren müssen in dieser Tabelle jeden genetisch veränderten Organismus nachhalten, und das über Jahre hinweg. Die für uns verantwortliche Bezirksregierung legt sich bei der Art der Dokumentation nicht fest - sie kann ausdrücklich auch in einem elektronischen Laborbuch geschehen.

1.2 Verankerung im Curriculum

Trotz der stetig wachsenden Benutzung besteht häufig eine interne Schwelle das elektronische Laborbuch auch einzusetzen. Die Motivation der Mitarbeitenden ist hier der wesentliche Punkt, an dem angesetzt werden muss. So sind wir vom ZIM bereits mit Forschenden im engen Austausch, das Thema schon früh Studierenden nahezubringen und attraktiv als Teil des Curriculums darzustellen, um damit die positive Haltung von Studierenden gegenüber Digitalisierungsprojekten auch bei der Digitalisierung von Laborarbeiten zu nutzen.

Mit dem Ziel einen Kulturwechsel hin zu elektronischen Laborbüchern zu gestalten, führte man, als Pilotprojekt in der Biologie, die Master-Studierenden des Kurses "Cellular and molecular analysis of brain function" mit Hilfe von Videotutorials in die Laborbuchsoftware ein, um unter Laborbedingungen erste Schritte in fluorometrischem Imaging zu erlernen, d.h. wie Daten generiert und ausgewertet werden sollen. In enger Zusammenarbeit mit den Lehrenden kann so das Laborbuch für die Lehrzwecke optimiert werden. So lernen junge Wissenschaftler früh den Umgang mit elektronischen Laborbüchern und werden in der Forschung an diese neuen Standards anknüpfen.

Durch die Einführung von ELB in der Lehre erhoffen wir uns eine schnellere und breitere Akzeptanz der elektronischen Laborbücher, um somit die Forschungsarbeit im Labor zu optimieren und die Digitalisierung dort voranzutreiben. Wir denken, dass ELB eine gute Möglichkeit sind bereits frühzeitig in der Lehre an Forschungstätigkeiten und Vorgehensweisen herangeführt zu werden und Potentiale für selbstständiges Lernen im Labor bieten.

2 Interne und externe Vernetzung

Um grundsätzliche Bedarfe von Forschenden zu identifizieren hat uns die Gründung einer NRW-weiten Arbeitsgruppe zum Thema ELB geholfen [EA22]. Die Kollaboration der Infrastruktureinrichtungen hat uns voneinander lernen lassen: So erhielten wir Hilfe bei der Implementierung von einer zentralen Authentifizierungsmethode, und helfen anderen beim Einsatz in der Lehre oder wenn es um konzeptionelle Fragestellungen geht. Unter anderem wird in dieser Arbeitsgruppe über die technische Einführung, sowie IT-Sicherheitskonzepte und Fragen bezüglich DSGVO und Personalratsbeteiligung diskutiert. Die Vernetzung der verschiedenen NRW-Einrichtungen bietet auch einen guten Überblick, denn neben eLabFTW verwenden diese Einrichtungen u.a. auch Labfolder, eLabJournal oder RSpace als Laborbuchlösungen [Fo22].

Wir erhielten einen besseren Überblick der Forschungsszenarien am eigenen Standort, indem wir verschiedene Forschende aus unterschiedlichen Fachbereichen eingeladen haben, um über ihre Anforderungen zu sprechen, und um sogenannte Use cases zu entwickeln [ZK19]. Aktuell treffen wir uns monatlich mit Power-Usern, um über

Probleme und Möglichkeiten in Bezug auf die Nutzung von elektronischen Laborbüchern an der HHU zu sprechen.

3 Stabilisierung des Service

Der Service eLabFTW ist aus einem Forschungsdatenmanagement-Projekt heraus erwachsen, und hat zum Status Quo immer noch einen gewissen Projektcharakter. Auch sind technische Tests nur das notwendige Kriterium, zusätzlich muss die Zielgruppe den Service auch aktiv nutzen wollen. Es wurden Gespräche mit dem Personalrat geführt, erste Versionen von Verfahrensverzeichnissen erstellt und über Datenschutzbetrachtungen nachgedacht. Man muss den Spagat zwischen den Datenschutzrechten der Einzelnen und der guten wissenschaftlichen Praxis meistern, also die Löschmöglichkeit der personenbezogenen Daten gegen Aufbewahrungs- und Nachweispflicht abwägen.

Wichtige Punkte, die bei einer Einführung von elektronischen Laborbüchern bereits vorher bedacht werden müssen, sind u.a.: Wer betreibt und verwaltet auf langfristige Sicht die Systeme? Wer ist Ansprechpartner für den Nutzenden? In welche Richtung soll der Dienst weiterentwickelt werden und welche Punkte des Forschungsdatenmanagements werden dadurch abgedeckt, welche nicht?

Dennoch war die produktive Testphase für uns sehr wichtig und nötig. Die Kreativität der Forschenden im Hinblick auf die Verwendung von Funktionen zu beobachten war nicht nur sehr interessant, sondern auch hilfreich bei der zu leistenden Überzeugungsarbeit.

4 Fazit

Die flexible und auch lizenzfreie Einsatzmöglichkeit der Open Source Labborbuchlösung eLabFTW bietet für uns breite Einsatzmöglichkeiten, sowohl in der Lehre mit sehr vielen Studierenden wie auch im individuelleren Forschungseinsatz mit Anforderungen an die Anpassbarkeit für Forscherteams und die Nutzung von Anwendungsschnittstellen. Offen bleibt für uns weiterhin die Frage, ob man mit dem bestehenden Dialog innerhalb der Nutzendengruppe weiterhin einen gemeinsamen monolithischen Plattform-Ansatz, durch die Verständigung auf gemeinsame Anforderungen, fortführen kann oder ob eine verteilte Strategie notwendig wird. Andere Einrichtungen teilen beispielsweise nach Fakultäten auf aber verlieren damit Synergien bei übergreifenden Projekten. Die Forschenden zu einer eLabFTW-Nutzendengruppe zusammen zu bringen, hat sich als sehr wertvoll erwiesen, um zu gemeinsamen Anforderungen zu kommen, aber auch um Neueinsteigende oder skeptische Nutzende mit konkreten Beispielen aus anderen Forschergruppen abzuholen.

Bibliography

[AL19] Adam, B.; Lindstädt, B.: Elektronische Laborbücher im Kontext von Forschungsdatenmanagement und guter wissenschaftlicher Praxis – ein Wegweiser für die Lebenswissenschaften. ZB MED, Köln, 2019.

[EA22] ELB.nrw AG zu Elektronischen Laborbüchern, https://wiki.hhu.de/x/vAMdB, accessed: 11/07/2022.

[Fo22] Forschungsdaten.org, https://www.forschungsdaten.org/index.php/Elektronische_Laborb%C3%BCcher, accessed: 11/07/2022.

[He20] Hewera, M. et al.: An inexpensive and easy-to-implement approach to a Quality Management System for an academic research lab. F1000Research, 9:660, 2020. https://doi.org/10.12688/f1000research.24494.2

[Kn20] Knipprath, N. et.al.: Schlussbericht FoDaKo - Forschungsdatenmanagement in Kooperation. Universitätsbibliothek der Universität Siegen, 2020. https://doi.org/10.25819/ubsi/728

[ZK19] Zulauf, B.; Knipprath, N.: Electronic Lab Notebooks - early research practice in teaching. In 12th International Conference of Education, Research and Innovation (ICERI2019), Seville, SPAIN. https://doi.org/10.5281/zenodo.4787258

A Snapshot of Essential IT-Related Challenges of Universities: A Literature Analysis

What drives IT in higher education today and tomorrow?

Stephan Leible [1], Max Ludzay [2]

Abstract: Universities have rarely been challenged to radically evolve and transform as they have been in the Covid-19 pandemic. The conduct of teaching and research, as well as the environment of staff and administration, have had to digitalize and transform under time pressure. This development has placed and continues to place high demands on IT and its infrastructure. In addition, new challenges have emerged, such as technology-based future skills, cybersecurity, and the need for an agile university-wide IT governance. Mastering these requires IT to address the present and the future already today. Therefore, in this article, we conducted a literature analysis to identify and examine current and future challenges for IT and its infrastructure in universities. We identified twelve overarching challenges, which we describe individually. This paper provides starting points for further IT-related research and essential viewing angles for IT governance in universities for practitioners.

Keywords: Digitalization, Challenges, Universities, IT Governance, Digital Infrastructure, Digital Education, Higher Education

1 Introduction

The IT and digital infrastructure of universities[3] are a cornerstone of operations for their stakeholders, as Gilch et al. illustrate in their digital model university [GJW21]. Among numerous factors, this includes the management of students and employees and goes beyond the provision of IT services such as learning management systems (e.g., Moodle), video conferencing systems (e.g., Zoom) to specific solutions for the implementation of new (digital) teaching concepts. What it would look like without IT and its infrastructure has been shown by the case of the University of Giessen, which we will briefly discuss later [Ma20]. Without information and communication technologies, everyday university life, especially during Covid-19, would be inconceivable [AHL20]. The IT manages and controls the digital infrastructure and has an increasingly prominent and crucial role in meeting stakeholders' current and future needs in higher education environments [WM17].

[1] Universität Hamburg, Fakultät für Mathematik, Informatik und Naturwissenschaften, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany, stephan.leible@uni-hamburg.de, 0000-0002-8742-4505
[2] Universität Hamburg, Fakultät für Mathematik, Informatik und Naturwissenschaften, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany, max.ludzay@uni-hamburg.de, 0000-0002-8805-9618
[3] We use the term “university” in this paper for all types of higher education institutions.

doi:10.18420/inf2022_117

One of the growing needs is the agile deployment and use of appropriately configured and secure software for the diverse IT technologies of the future, which are playing a role in more and more disciplines [KM21]. Therefore, the topic of digital and data literacy and its interdisciplinary teaching is a requirement arising from global development [Ri15], [RS20]. Using data as a risk-mitigating element for forecasting and decision support or as the base for reasoning to explain phenomena plays a fundamental role in all disciplines [RS20]. Over the past two years, the Covid-19 pandemic has led as an accelerator for the digitalization of teaching concepts, operational work, and communication in universities. Thus, as one example, Massive Open Online Courses (MOOCs) or distinct divisions of teaching into synchronous (live) and asynchronous parts can be found more and more frequently [KL19]. In the asynchronous part, the lecturers prerecord the classical course content via videos so that they can be accessed by the students on-demand. The synchronous parts are used instead for exercises and direct exchange in the form of discussions about the content. These teaching concepts need to be enabled by universities' IT and bring their own challenges for the provided platforms, tools, and data management [AHL20]. In addition, the flexible integration of new software in compliance with general and university-specific data protection regulations is becoming increasingly important to enable contemporary teaching with modern applications. IT can also provide beneficial services for the researchers themselves and research data management (RDM) [Au19].

Universities as expert organizations have a unique structure compared to other types of organizations, be they private entities or public authorities [RBS15]. This comes through autonomously acting faculties with experts in different disciplines and domains operating in the sphere of action, centralized by a governing body of the university, the presidium. Since these are traditionally grown organizations, IT is characterized by its demand-oriented development process. Complicating matters are inertia and multi-layered bureaucratic processes embedded in the IT governance of universities that constrain agile ways of working. Flexibility, however, is necessary to teach students in current technologies and the latest software and hardware for high-quality academic education. For countering and mastering current challenges, for example, frameworks such as COBIT [Ge20] are applied in universities, data and digital strategies are developed, and C-level positions such as Chief Information Officer (CIO) and Chief Digital Officer (CDO) are filled to spearhead IT governance and digital development [Ho15].

This paper captures current and future challenges for higher education IT based on the literature analyzed. The literature analysis was conducted using the five phases defined by Fettke [Fe06] (1) problem formulation, (2) literature search, (3) literature review, (4) analysis and interpretation, and (5) presentation. The *(1) problem formulation* is described via the following research question (RQ): *What are the current and future challenges facing university IT and its digital infrastructure?*

The *(2) literature search* was primarily executed using Google Scholar as a literature search engine with the inclusion of various databases with the search term (("IT Challenges" OR "Challenges") AND ("Universities" OR "Higher Education")). The search was limited to results with a maximum age of 10 years, i.e., 2012, assuming that

older literature no longer represents the current state of challenges and problems. After several phases of analysis with the exclusion of duplicates, the utilization of the titles, the abstracts, and finally, a full-text analysis, as well as a forward and backward search as recommended by Webster and Watson [WW02], 27 paper relevant to the problem formulation remained. The *(3) literature evaluation* is performed in section 2 (Current and Future IT Challenges of Universities), and the *(4) analysis and interpretation* are performed in section 3 (Discussion and Limitations), which also contains the limitations of this paper. Section 4 (Conclusion) provides the conclusion. This article, in its entirety, represents *(5) presentation* according to Fettke's five phases [Fe06].

2 Current and Future IT Challenges of Universities

All challenges identified are shown in Tab. 1. In this chapter, we will describe each challenge with its specifics. We want to note that many of the challenges are related. However, these cross-connections have not been formulated in detail, as the respective infrastructures of the universities are different, and the relationships are therefore not universally valid.

#	Challenge	Literature Findings
Ch1	Cybersecurity	[Bo17], [Cu19], [Ma20], [UW21]
Ch2	(Interdisciplinary) future skills	[Ha21], [He21], [GJW21], [KLS17], [KM21]
Ch3	Complexity of infrastructure	[Bo17], [FF18], [He22], [Ho15], [Ro21]
Ch4	Mobilization	[AHL20], [Bo17], [GJW21], [KM21], [vv18]
Ch5	Scalability	[GJW21], [Qa19], [WM17]
Ch6	(IT) staffing & shortage of skills	[GJW21], [PW17]
Ch7	Data-driven culture	[Da15], [TZS22]
Ch8	Data & IT governance	[AHL20], [Bo17], [Da15], [Ge20]
Ch9	Agility for change & innovation	[KM21], [PK16], [He21], [MS21], [MS22]
Ch10	Outdated infrastructure	[Ch18], [PS20], [WM17], [ZM21]
Ch11	New learning models	[Gi17], [GJW21], [He21], [Jo20], [KLS17], [KM21], [UC22]
Ch12	(Top) management commitment	[Ge20], [GJW21], [Ho15], [WM17]

Tab. 1: IT Challenges that universities (will) face

Like every organization, universities must protect the data of their users. Given the amount of sensitive data such as student and employee payment information, personnel data, research data not yet intended for publication, teaching documents, and many others [UW21], the issue of **cybersecurity (Ch1)** is particularly relevant [Bo17]. To protect personal data, the European Union has had, among other things, the General Data Protection Regulation (GDPR) since 2018 [Eu18]. However, compliance with these regulations is a challenge in their implementation to create a resilient digital infrastructure. A complex tension exists between compliance with data protection and using modern software, where data is often exchanged, for example, via data notes in foreign countries. In addition, human factors can also be a safety threat [Cu19]. Cases like those in Germany at the University of Giessen in 2019/2020 must be avoided [Ma20]. A successful hacker attack forced the university to take most of its digital services offline for about a month. The effects were far-reaching: Seminars and exams were canceled, for some students, this involuntarily prolonged their studies, and due to the lack of digital interfaces, alternative analog processes had to be used to improvise [Ma20]. This incident clearly shows an inseparable dependency on the digital infrastructure for the orderly flow of university processes and that this must be protected as a result. In the future, this dependency will be even more pronounced than now, as the digitization of processes and data will continue to increase, and the Covid-19 pandemic has further accelerated the process [UW21].

There is a steadily growing need for cross-cutting competencies such as **digital, data, and other skills (Ch2)** that will greatly benefit business and civil society in the future [Ha21]. In exercising their educational mission to prepare and train academics and young scientists, universities must respond to these requirements [He21], [KM21]. The embedding of content becomes challenging if it is taught as interdisciplinary skills across all faculties, such as digital and data literacy [GJW21]. The addressees are primarily students but also lecturers who integrate such competencies sustainably into their courses [KLS17]. An agile centralized or decentralized IT organization with comprehensively defined responsibilities is conducive to the digital infrastructure of such efforts. Particularly in the case of decentralized, faculty-centric IT units, the question often arises in the context of collaborations as to which team provides the services and is responsible for them.

The development of digital innovations is advancing every day, and so is the number of tools, services, processes, and interfaces increasing the **complexity of the surrounding infrastructure (Ch3)** [Bo17]. For example, the pandemic has driven the proliferation of various video conferencing systems, remote work environments, and digital learning tools. Emerging demands from new or further diversifying disciplines also add to the complexity. Efficiently and effectively managing these numerous and diverse IT infrastructure components requires careful planning and coordination through IT governance since the available resources are usually very limited [He22], [Ho15]. An aggravating circumstance is the IT that has traditionally grown up at universities. Thus, the IT and its digital infrastructure are not holistically designed from the ground up to function in a resource-efficient manner but have evolved incrementally through management direction, regulations, or individual needs without structured coordination of

the big picture and its individual components. Thus, in three university digital ecosystems, we observed numerous obsolete, unused, redundant, forgotten, and inefficient software that complicates their IT environment.

A phenomenon worth mentioning that arises in many kinds of organizations but particularly in higher education environments is the formation of silos [Ro21]. In these, processes, content, technologies, tools and data may be hoarded and only used for the organizationally specific goals, for example, of a faculty or department, fostering tribalism [FF18]. For instance, the same problem in dealing with students or in the context of digitalization can occur in several organizational areas, and each of the areas plans, develops, and integrates its own solution and doesn't share its knowledge, while one approach would have sufficed [FF18]. It would be desirable to develop and implement joint solutions that are inclusive and, above all, holistically viewed in a more resource-efficient manner. This is a major challenge for universities and their (IT) architecture management.

Another aspect that became necessary during the Covid 19 pandemic and will take on an essential role in the post-Covid 19 era is the **mobilization of the digital infrastructure and its services (Ch4)** [Bo17]. Meanwhile, events on campus are once again taking place more frequently and regularly. Nevertheless, the development shows that fully digital or hybrid forms of teaching have become established and remain firmly anchored in the repertoire of universities [GJW21]. This results in integrating various services for digital education in the IT infrastructures. In addition to video conferencing systems, these include (collaborative) tools such as online whiteboards, design and programming environments, and computing capacities. Students want digital access to education, but they also want to carry out administrative activities with their university in the digital space in order to save the time that would have been needed to travel to campus [AHL20]. Employees are becoming accustomed to home office as a working model and wonder whether the daily trip to the university is even necessary when they can do most activities at home without restrictions [KM21]. In this context, new-work initiatives are an issue [vv18]. These include, for example, room booking systems that allow on-site workplaces to be shared by several employees [vv18]. Such new services must, in turn, be integrated into the IT infrastructure, and the interacting individuals must become accustomed to the new forms and environments of work over time. In the competitive job market, the mobility characteristic is proving to be an attractive feature. We see the development that the digital and physical campuses will coexist in the future and can be used individually by each student as they see fit. However, universities are places where students come together, engage in academic discourse, and acquire essential competencies, which include social skills. The purely digital role should be reserved for distance universities designed for that purpose. Further cultural reflection on how much digital is too much is an ongoing interdisciplinary issue.

IT structures are typically subject to the issue of **scalability (Ch5)**, especially if the growth is planned and it is possible to be unexpectedly rapid. Ensuring scalability primarily depends on how many resources are available in the (technical) infrastructure [GJW21].

However, this is also related to the effective and efficient use of available resources. For example, designing a process well can mean massive differences in scalability as the activity runs faster, freeing up resources for other activities [WM17]. The scarcity of financial resources and associated cost pressures present universities with a non-trivial task in this regard. One approach that is becoming increasingly popular is the outsourcing of individual or multiple services to external cloud systems [Qa19]. This transfers some of the responsibility to a service provider. Once again, this must be planned in detail, be feasible with the available financial resources, and be permissible from a regulatory point of view. Scaling also affects education itself, such as the delivery of exercises in large courses that require IT support to be feasible. We expect the number of university students and employees to grow steadily in the coming years, making scaling education as well as IT and its digital infrastructures an ongoing and important topic.

The **demand for (skilled) IT staff (Ch6)** on the job market is already very high, as evidenced by many unfilled positions and a rising salary level to address the low supply of (sufficiently trained and experienced) workers [PW17]. Demand is even expected to increase as digitalization continues. Since the private and public sectors, including universities, operate in the same job market, they compete for the supply of IT professionals [GJW21]. In Germany, university jobs are generally subject to collective bargaining agreements, which is not necessarily negative. However, salaries and other incentives often cannot compete with those in the private sector [GJW21]. In addition to recruiting new professionals, it is also a matter of training existing ones to meet current and future skill needs and perform their jobs [GJW21]. The IT sector is fast-moving, and new software products and technologies are added in ever shorter cycles, whose integration and application must be learned and trained. If there is too little personnel capacity or knowledge in the organization, external consulting or IT service providers are often the only alternatives. Overall, it will not be any easier for universities to acquire or train the IT professionals they need in the future.

A growing number of organizations are **adopting a data-driven culture (Ch7)** to improve their performance and decision quality [TZS22]. They all generate verifiable data in various forms and types that have great potential to understand an organization, forecast the future, measure what processes and actions are effective, or whether a product is successful. Because data is ubiquitous in digital environments, a tremendous number of applications exist to analyze them. Universities generate vast amounts of data in each of their pillars, be it education, research, administration, or transfer [Da15]. However, the data is rarely analyzed. Therefore, investment in adequate business intelligence tools, AI, and the teaching of appropriate data literacy is necessary [Da15], [TZS22]. A data-driven culture is not primarily about using data as the sole basis for decisions but about backing them up with information to argue for a direction. This is compounded by the complex challenge of creating employee awareness of data, how to interpret it, and how to deal with it.

Creating a data-driven culture requires a clear framework in the form of **data and IT governance (Ch8)**, which includes processes, guidelines, and standards to define the

organizational goals and support their implementation. Various frameworks such as COBIT and ITIL exist for this purpose, but applying them to the university context is not an easy endeavor [Ge20]. Creating and actively using data is a general challenge in the coming years. It is becoming particularly crucial because, for example, IT risks and data security issues are interwoven with it [Ge20]. In addition, there are the questions of (legally) correct and sustainable data storage, data management, and a uniform specification of integration processes of new data into the existing data ecosystem [Bo17], [AHL20]. Pre-definition for integrating new data is important to ensure that computer-aided analysis and evaluation are possible. Besides the potential applications for improving the university for students and employees, data can also serve as a foundation for argumentation and communication to discuss and support hypotheses, theories, and evaluations in research [Da15]. In the case of scientific data, the FAIR Guiding Principles were established in 2016 [Wi16]. FAIR stands for: Findable, accessible, interoperable, and reusable digital assets. Creating data governance and strategy and gaining its acceptance across the university is a complex, multi-year endeavor. It requires the presence of appropriate professionals and the will of all to realize it successfully. It becomes particularly complicated when there are separate IT units, for example, in the faculties, which act autonomously.

The topics of **agility for change and innovation (Ch9)** must be addressed in universities at all levels in order to be able to demonstrate sufficient flexibility, efficiency, and speed of adoption in the future [He21], [KM21], [PK16]. The entire organizational (IT) architecture must be coordinated for this, which makes this challenge particularly large and complex [MS21]. For expert organizations with autonomously acting working groups, achieving or initiating change is a multi-faceted task. Based on our observations, some IT architectures hinder agility due to an unfavorable mix of multi-tier centralization and decentralization. An example of this would be the following three-layer structure example from practice:

1. Centralized organization of the critical IT infrastructure by a central computer center,
2. Decentralized (agile) units per faculty with smaller teams to serve their individual needs in the faculties, and
3. Decentralized temporary project teams (mostly financed by third-party funds) that fulfill the respective project requirements.

In the case of internal faculty projects, the communication chain with this structure is not optimal, but it is manageable, and responsibilities are regulated. Recently, however, multi-faculty projects such as the teaching of cross-cutting competencies have become increasingly common [MS22]. In this context, there may be difficulties in transferring tasks, as the second level sees itself as responsible only for its faculty. In addition, coordination and documentation among these units are fundamental to governing which services are provided at which level by which unit. The project teams on the third level are usually only temporary for about three years and act autonomously. However, they must communicate with the first and second levels and build on their services to avoid

wasting resources and operate sustainably.

The practical example is one variant of many possible and exemplifies how many parameters and organizational intersections play a role and influence agility, change, and thus innovation [MS21]. Especially due to the silo situation in expert organizations, the democratization of information also takes on an essential function in universities for agility and internal as well as external collaboration. Weaknesses of the described three-layer structure can be lengthy processes, unclear responsibilities, inefficient coordination and use of resources, and slow change overall. As is common in large, established organizations, the first steps of change should be taken on a smaller scale in individual areas rather than initiating an entire transformation at once, if possible. The goal must be to allocate the available, limited resources to the internal and external needs and impulses in the best possible way and to adapt to the operational environment in a minimum of time [MS21].

Another challenge currently and in the future is handling the **outdated systems and software in the IT infrastructure (Ch10)** of universities, especially in developing countries [PS20], [ZM21]. Often, these are deeply anchored in one or more faculties, for example, when it comes to content, campus, and learning management systems. Lecturers have become accustomed to these programs and adapted their teaching scenarios to them. However, the handling is often uncomfortable, the user interface is no longer up-to-date and presentable, and functions such as data analysis and visualization are missing. In addition, such systems often have high maintenance requirements and are vulnerable to cyber-attacks [Ch18]. Other components to consider are old systems that are no longer used and are placed as corpses in the digital infrastructure. Depending on the complexity and rootedness of the systems, adaptations here are less or very costly. The larger the user base, the more likely the university's management, at the top of the hierarchy, will have to be involved in initiating and implementing a change to update the infrastructure, assuming the necessary resources are available [WM17].

One major transformation was the development of **new learning models (Ch11)** needed in the Covid-19 pandemic in an all-around digital learning environment. This change has required a lot from students and lecturers, depending on their existing digital skills [Gi17]. Technical and didactic elements together play here an essential role in ensuring the experience as well as the learning success of and interaction with students [Gi17]. New tools were needed for the IT infrastructure to execute distance learning, which we have already mentioned with video conferencing systems and collaborative learning environments [KLS17]. However, we see this development as far from complete, as such models are just becoming established and entrenched in a post-Corona era [GJW21], [He21], [KM21]. But it's not just distance learning that's changing; on-campus education also shows new facets. Dualization into asynchronous prerecorded teaching videos mixed with synchronous physical teaching units in which (extending) exercises and discussions to the asynchronous parts take place, show up frequently. New tools are becoming more prevalent such as Jupyter Notebooks as an interactive, collaborative environment used for individual exercises or entire courses, as Berkeley University has demonstrated for several

years with its Data 8 cross-sectional introductory data science course [UC22]. In the case of Jupyter Notebooks, common elements such as text, images, and videos can be interactively combined with mathematics, statistics, or programming in the notebook [Jo20]. The computations are usually done on the university's servers to keep the use of the software as low-threshold as possible [Jo20]. The learning models and environments as well as tools mentioned are only an exemplary glimpse of the future developments that must be provided and supported on the part of IT structures and thus one of the current and future challenges for university IT.

A characteristic accompanying almost all significant (successful) changes is the **commitment of (top) management (Ch12)**. Problems arise if there is no designated C-management position for digitalization issues, such as a CDO or CIO, who can drive forward the university's digital development with the appropriate experience, qualification, responsibility, and resource capacity [Ge20], [GJW21], [Ho15]. Such management positions can be justified by the far-reaching and all-pervading digitalization, for which a superordinate control authority is beneficial. In some cases, the overall tasks a CDO or CIO should do in full time are integrated as an additional function, e.g., into the repertoire of the rector, who, on the one hand, usually does not have enough time for it and, on the other hand, often does not have the appropriate qualification background for this functional area [WM17]. Another aspect of the CDO and CIO positions is to regulate clear responsibilities and grant them the necessary hierarchical authority so that binding goals and tasks, as well as resources, can be delegated [Ge20], [Ho15]. Many of the challenges described cannot be mastered without a (top) management commitment and impetus, especially when a large number of stakeholders need to be motivated.

3 Discussion and Limitations

Within a funded project that aims to foster digital and data literacy as cross-cutting competencies across the university among students and lecturers, we confirmingly encountered, to varying degrees, all twelve challenges identified from the literature. In that project, we use employee-driven digital innovations as a basis, i.e., innovations in the form of new and expanded digital or data literacy-promoting teaching concepts and courses designed, developed, and provided by the internal university teaching employees [LLN21], [Op22]. Internally, a funding pot is made available for this purpose as crowdfunding for innovative teaching endeavors [Jo17]. The special note in this project results from the collaboration of all faculties and the integration of new competencies, related technologies, and IT services. For example, in the project, the question of which of the IT teams is responsible frequently arose when tasks or problems affected multiple faculties. We have also encountered the situation where one of the faculties has a working digital solution to a challenge but does not want to share it with other organizational units to save internal resources. Therefore, the other faculties must develop their own solutions unless a management directive triggers the sharing of the solution. This case represents the epitome of a silo solution, where the various stakeholders pursue the goals of their

organizational unit or department but not those of the university as a whole. The illustration of our experience is intended to be an exemplary view of several aspects that should be considered when designing an IT governance and organizational structure.

The status quo of universities for the years ahead is currently just emerging, so more IT-related challenges may arise. However, we are confident that the twelve mentioned in this paper will occupy IT at universities in the coming years and, in some cases, in the next decade. More than ever, a look should be taken at the structures of benchmark companies on the topics of agility, IT governance, and adaptation of innovations, among others. We need to start shaping the university a decade from now. Every concept and implementation step already contributes to this and anticipating the future. Although a decade is long, the upcoming (far-reaching) changes need that time, especially in universities.

This work is not without limitations. Due to the limitation of the literature to Google Scholar and the databases linked therein, we do not claim complete coverage of all relevant literature. Furthermore, interesting literature may not have been found due to the definition of the overall simple search term without additional synonyms. Moreover, our observations and experiences, as well as most of the literature examined, refer to the German area. Altogether, universities represent individual organizations that are structured and function differently in detail, so the challenges described may vary in size, small or large, depending on the initial situation. They also differ further in terms of the type of university. Nevertheless, a consideration of each type would have gone beyond the scope of this article. However, despite the limitations, the results of our analysis should answer the RQ defined at the beginning about current and future challenges in university IT and provide promising starting points for research and perspectives for practice.

4 Conclusion

This paper presents and describes twelve current and future challenges for IT in universities and their digital infrastructure identified from the literature. After a long period of suffering for society in health and social aspects due to the Covid-19 pandemic, universities are slowly resuming their regular operations. However, these are not returning as we knew them before the pandemic [KM21]. New digital ways of teaching, virtual environments for students and employees in the form of new work initiatives, and many other aspects remain in place with varying intensity depending on the university [KM21]. If we reflect on this time, we can undoubtedly find positive elements in the accelerated digital development of universities, but numerous processes and tools had to be integrated into the digital infrastructures under great time pressure, so planning has been noticeably neglected in many places. Each of the challenges outlined in this paper offers starting points for further research and an impetus for active discourse at the scientific and practitioner levels about how these obstacles can be overcome in the coming years. An exciting addition to our work would be to conduct a qualitative analysis in the form of interviews to expand the challenges or find new ones that have yet to be considered.

5 Bibliography

[AHL20] Auth, G.; Heyde, M. von der; Lucke, U.: Hochschule 2030 - Welche Entwicklungen der IT prägen die Zukunft der Hochschulen. In (Reussner, R.; Koziolek, A.; Heinreich, R. Eds.): Informatik 2020, pp. 449–451, 2020.

[Au19] Auth, G. et al.: Digitalisierung des Forschungsprozesses aus Sicht von Forschenden – durch Serviceintegration zum persönlichen Forschungsinformationssystem. In (Barton, T.; Müller, Christina, Seel, Christian Eds.): Hochschulen in Zeiten der Digitalisierung. Lehre, Forschung und Organisation. Springer Fachmedien Wiesbaden, pp. 287–307, 2019.

[Bo17] Bojinov, B.: Problems IT Management in Universities. SSRN Electronic Journal, 2017.

[Ch18] Christopher R. Harrell et al.: Vulnerability Assessment, Remediation, and Automated Reporting: Case Studies of Higher Education Institutions: Proceedings of the 2018 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 148–153, 2018.

[Cu19] Cuchta, T. et al.: Human Risk Factors in Cybersecurity. In (Goda, B. et al. Eds.): Proceedings of the 20th Annual SIG Conference on Information Technology Education. ACM, New York, NY, USA, pp. 87–92, 2019.

[Da15] Daniel, B.: Big Data and analytics in higher education: Opportunities and challenges. British Journal of Educational Technology 5/46, pp. 904–920, 2015.

[Eu18] European Union: General Data Protection Regulation (GDPR). https://gdpr.eu/tag/gdpr/, accessed 12 Jul 2022.

[Fe06] Fettke, P.: State-of-the-Art des State-of-the-Art. Eine Untersuchung der Forschungsmethode 'Review' innerhalb der Wirtschaftsinformatik. Wirtschaftsinformatik 4, pp. 257–266, 2006.

[FF18] Friedman, H.; Friedman, L. W.: Does Growing the Number of Academic Departments improve the Quality of Higher Education? Psychosociological Issues in Human Resource Management 1/6, pp. 96–114, 2018.

[Ge20] Gerl, A. et al.: Applying COBIT 2019 to IT Governance in Higher Education. Establishing IT governance for the collaboration of all universities and universities ofapplied sciences in Bavaria. In (Reussner, R.; Koziolek, A.; Heinreich, R. Eds.): Informatik 2020, 517-530, 2020.

[Gi17] Gillett-Swan, J.: The Challenges of Online Learning: Supporting and Engaging the Isolated Learner. Journal of Learning Design 1/10, p. 20, 2017.

[GJW21] Gilch, H.; Jungermann, I.; Wannemacher, K.: Modellierung einer digitalen Hochschule post-Corona. Einflüsse und Nachwirkungen der pandemiebedingten Transformation. In (Gesellschaft für Informatik Ed.): Informatik 2021, pp. 1713–1725, 2021.

[Ha21] Hachmeister, N. et al.: Balancing Plurality and Educational Essence: Higher Education Between Data-Competent Professionals and Data Self-Empowered

Citizens. Data 2/6, p. 10, 2021.

[He21] Heyde, M. von der et al.: Vision, Mission und Governance der Landesstrategien im Zeitalter der Hochschul-Digitalisierung. Unterschiede und Gemeinsamkeiten im Vergleich von fünf Bundesländern in Deutschland. In (Gesellschaft für Informatik Ed.): Informatik 2021, pp. 1675–1687, 2021.

[He22] Heyde, M. von der et al.: Gemeinsame Digitalisierung durch die Entwicklung von IT-Governance-Strukturen auf Landesebene. Wirtschaftsinformatik & Management, 2022.

[Ho15] Hotzel, H. et al.: IT Governance – role of a CIO in German Universities – a Survey by ZKI. PIK - Praxis der Informationsverarbeitung und Kommunikation 3-4/38, pp. 121–126, 2015.

[Jo17] Jovanovic, T. et al.: The Crowdfunding Idea Contest of BMW: Proceedings of the International Association for Management of Technology (IAMOT 2017), pp. 1–15, 2017.

[Jo20] Johnson, J. W.: Benefits and Pitfalls of Jupyter Notebooks in the Classroom. In (Khazanchi, D. et al. Eds.): Proceedings of the 21st Annual Conference on Information Technology Education. ACM, New York, NY, USA, pp. 32–37, 2020.

[KL19] Kirchner, K.; Lemke, C.: MOOCs als disruptive Innovation für die akademische Bildung. In (Barton, T.; Müller, Christina, Seel, Christian Eds.): Hochschulen in Zeiten der Digitalisierung. Lehre, Forschung und Organisation. Springer Fachmedien Wiesbaden, pp. 239–263, 2019.

[KLS17] Kebritchi, M.; Lipschuetz, A.; Santiague, L.: Issues and Challenges for Teaching Successful Online Courses in Higher Education: A Literature Review. Journal of Educational Technology Systems 1/46, pp. 4–29, 2017.

[KM21] Klyachko, T.; Mau, V.: Modern Challenges for Universities. In (van't Land, H.; Corcoran, A.; Iancu, D.-C. Eds.): The Promise of Higher Education. Springer Cham, pp. 233–237, 2021.

[LLN21] Leible, S.; Ludzay, M.; Nüttgens, M.: Ein IT-gestützter Innovationsprozess in der öffentlichen Verwaltung: Rahmenkonzept, Ideenmanagementsysteme und Online-Whiteboards. HMD Praxis der Wirtschaftsinformatik 5/58, pp. 1108–1128, 2021.

[Ma20] Mahrt, K.: Wenn die Uni plötzlich offline ist - Hackerangriff auf die Universität Gießen. https://hochschulforumdigitalisierung.de/de/blog/Hackerangriff-Universitaet-Giessen, accessed 12 Jul 2022.

[MS21] Menon, S.; Suresh, M.: Factors influencing organizational agility in higher education. Benchmarking: An International Journal 1/28, pp. 307–332, 2021.

[MS22] Menon, S.; Suresh, M.: Enablers of technology agility in higher education. The International Journal of Information and Learning Technology 2/39, pp. 166–196, 2022.

[Op22] Opland, L. E. et al.: Employee-driven digital innovation: A systematic review and a research agenda. Journal of Business Research 143, pp. 255–271, 2022.

[PK16] Pucciarelli, F.; Kaplan, A.: Competition and strategy in higher education: Managing

complexity and uncertainty. Business Horizons 3/59, pp. 311–320, 2016.

[PS20] Pavlin, S.; Sušanj, Z.: The academic profession in the Western Balkans: Time for a turning point? European Journal of Education 4/55, pp. 471–475, 2020.

[PW17] PWC: Fachkräftemangel im öffentlichen Dienst, 2017.

[Qa19] Qasem, Y. A. M. et al.: Cloud Computing Adoption in Higher Education Institutions: A Systematic Review. IEEE Access 7, 2019.

[RBS15] Rybnicek, R.; Bergner, S.; Suk, K.: Führung in Expertenorganisationen. In (van Dick, R.; Felfe, J. Eds.): Handbuch Mitarbeiterführung: Wirtschaftspsychologisches Praxiswissen für Fach- und Führungskräfte. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–11, 2015.

[Ri15] Ridsdale, C. et al.: Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report. SSHRC, 2015.

[Ro21] Roper, L.: Encouraging Interdisciplinary Collaboration: A Study of Enablers and Inhibitors Across Silos in Higher Education. Interdisciplinary Journal of Partnership Studies 1/8, p. 6, 2021.

[RS20] Raffaghelli, J. E.; Stewart, B.: Centering complexity in 'educators' data literacy' to support future practices in faculty development: a systematic review of the literature. Teaching in Higher Education 4/25, pp. 435–455, 2020.

[TZS22] Teng, Y.; Zhang, J.; Sun, T.: Data-driven decision-making model based on artificial intelligence in higher education system of colleges and universities. Expert Systems, 2022.

[UC22] UC Berkeley: Data 8. http://data8.org/, accessed 12 Jul 2022.

[UW21] Ulven, J. B.; Wangen, G.: A Systematic Review of Cybersecurity Risks in Higher Education. Future Internet 2/13, p. 39, 2021.

[vv18] van Marrewijk, A.; van den Ende, L.: Changing academic work places: the introduction of open-plan offices in universities. Journal of Organizational Change Management 5/31, pp. 1119–1137, 2018.

[Wi16] Wilkinson, M. D. et al.: The FAIR Guiding Principles for scientific data management and stewardship. Scientific data 3, 2016.

[WM17] Wimmer; Martin: IT-Governance an Hochschulen. Notwendigkeit, Stand und Wege zum Erfolg. Die Hochschule: Journal für Wissenschaft und Bildung 1/26, pp. 70–82, 2017.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 2/26, pp. xiii–xxiii, 2002.

[ZM21] Zarei, S.; Mohammadi, S.: Challenges of higher education related to e-learning in developing countries during COVID-19 spread: a review of the perspectives of students, instructors, policymakers, and ICT experts. Environmental science and pollution research, 2021.

What's Digital? Factors of perceived Digital Transformation of Higher Education in Germany

Correlations to Educause's Dx Self-Assessment

Markus von der Heyde [1]

Abstract: Higher education (HE) institutions in Germany are driven to address chances and risks, as well as methods to guide digital changes. Naturally, all leaders set their own agenda and define within their responsibilities an individual meaning of "digital" as "the ideal mix of useful, safe, practical and affordable" digital tools and processes. Do they all share a common understanding, and what is the core of this understanding in the context of HE? Based on the dataset of a national survey, the factors of Educause's Self-Assessment questionnaire on the Digital Transformation (Dx) are evaluated. A quantitative analysis of responses from one third of all universities in Germany was conducted to focus on the potential of the various topics and extract systematic patterns: Which group of topics or factors offer a potential to observe the change? Which factors might have the potential to drive the change? In addition, which factors appear not to have sufficient influence on the change and are not modulated by the perceived digital changes in HE? The four most prominent factors measured by effect size concern the topics "dealing with change", "flexibility of staff deployment", "new jobs and functions", and "importance of soft skills". Those topics are expected to gain further attention in the strategic discussions among leaders in higher education.

Keywords: Digital Transformation, Maturity Rating, Self-Assessment, Higher Education.

1 Introduction

Many work environments in universities have been changed and are still changing due to the use of digital tools and processes. Especially in areas where information technology (IT) is not the only relevant perspective, "Digitalisation" and related terms are often bundled into a broader understanding. Therefore, many factors are expected to be relevant for the digital transformation (Dx). The nature of the digital change in society and technology has been researched in various disciplines [We15, Ab15, UA17]. Many scholars have looked at the change itself in terms of chances, risks and why and how both work and are interconnected with our everyday lives.

Specifically in higher education (HE) the terminology around the levels of digital change has been reviewed [KHV+13, He17, Or20, Gr21]. In addition, a group of experts at

[1] vdH-IT, 99425 Weimar, Germany, info@vdh-it.de, https://orcid.org/0000-0002-6026-082X

doi:10.18420/inf2022_118

Educause[2] collected potentially relevant dimensions of the digital change and characterized each in three distinct statements. This collection was published as a Self-Assessment tool to guide institutions in HE in order to question and prepare the institutions and the main leadership groups for what they called the digital journey [Re20]. Educause as an organization suggests thinking of the digital transformation as "a series of deep and coordinated culture, workforce, and technology shifts that enable new educational and operating models and transform an institution's business model, strategic directions, and value proposition" [Ed21]. Educause therefore recommends discussing the digital change of the institutions at various organizational levels from multiple perspectives, in order to gain a complete picture and address dimensions which otherwise might only be implicitly covered or forgotten.

Focussing on many factors at once is rich and powerful but daunting, due to the complexity of multi-dimensional observations and control. Therefore, it is important to understand the nature of "digital" within the specific context of HE and the relative contributions of the different factors. Which of the factors are relevant in assessing the status of the digital journey – which are therefore indicators of the ongoing changes? Which other factors are potentially suited to drive or change the direction of the developments?

The specific status of digital change in German HE institutions was assessed in February 2022 in a national survey addressing the leadership level (CEOs, CIOs, CDOs, VPs for Digitalization, CTOs). This multiple-role-based approach was chosen due to the broad understanding of the roles in HE, which had been previously collected in two studies [AH22, HG22]. Of the 423 HE institutions in Germany, about one third responded to the survey and contributed valuable insights [He22a].

This article re-evaluates the dataset of this national survey [He22b] for the collected ratings on an idealized digitalization score based on the present status and the expected development in five years' time, as well as most of the statements of the Educause Self-Assessment (ESA) questionnaire.

The following research questions (RQ) are addressed by the statistical analysis and following discussion:

1. Which factors have been assumed to be relevant, but show no correlation to the current or expected future state of the universities?
2. Which factors contribute to the overall perceived state of digitalization of typical higher education institutions?

The remainder of this article begins with a short introduction to the methods. All relevant statistical results concerning the RQ are summarized and structured along with the ESA factors in the results and the appendix. The results are grouped for their qualitative similarity and discussed afterwards in depth in order to propose answers to the RQ. Further literature is used to put the findings into a context and provide relevant suggestions for

[2] Educause is a not-for-profit organization of HE institutions – see: https://www.educause.edu/about/

further research and open questions. The summary condenses the most important findings into a brief review of the current situation and the reception of the digital change in HE institutions in Germany.

2 Methods

A national survey on the current state of digital transformation of higher education in Germany was sent to participants at the CxO level of universities. In addition to the majority of the questions in the Educause Self-Assessment (ESA) of Dx[3], they were asked two questions:

- "Imagine an institution that has achieved *perfect* use of digital technologies and tools in all areas. How do you currently rate your institution compared to this imagined institution?"
- "Assess your institution in five years. How will your institution then compare to the *perfect* institution presented?"

Both statements were rated on a scale of "very different" to "quite similar" indicated by a continuous slider. The slider position corresponded to data values between 1-99, which were recorded. To clarify the use of "perfect", we offered the definition in this context as "the ideal mix of useful, safe, practical and affordable".

The idea to ask for these judgements is based on Educause' Dx rating conducted in [Ed21]. This method was adopted since the original questions contained a positive connotation about the effects of digital transformation. The chosen approach focusses on a neutral and personal perspective without addressing the consequences.

The various dimensions (also referred to as factors) which each were characterized in the ESA in three distinct statements were translated into German and also adapted to a minor degree to the German context (e.g., to match common terms). A group of experts provided feedback on draft versions of the questionnaire and its adaptation to the Germany context. In this process, the number of questions was reduced by excluding three questions (#3, #12, #22) which were seen as too specific for the target audience of the survey. The question on procurement of IT (#30) was complemented by the perspective of centralized vs. decentralized IT-administration. All remaining items of the Educause Self-Assessment were presented for a rating along the scale of all three statements per factor. If participants felt that the items should not be rated on one scale, they were presented with the three statements to be rated with a scale of "disagree - agree". Only a few participants made use of this more specific rating.

Due to the general approach of the overall survey, participants had initial choices to reduce their individual workload. Participants rated the amount of time they had to invest in

[3] See details about the Dx journey and the questionnaire.

participation, and their individual estimate of their own competency in the areas of leadership, digital transformation, and IT-administration and -operation. Based on these initial ratings, the questionnaire presented a subset of the questions (see [He22b] for a detailed description), aiming for a duration between 15 and 60 minutes. In consequence, the response rate for some items on the ESA was lower than on other items.

The collected data was controlled for quality. The ordinal data between 1 and 99 often did not comply with the assumption of normal distributions. Ratings were therefore divided in two groups by use of the median within each criterion, and both groups were compared in Mann-Whitney U-tests (=Wilcoxon-Mann-Whitney tests) with two-sided hypotheses without a presumption of direction of any effects. Results are reported for institutions rather than individuals by means of number of samples [n], median [m], power [β], effect size [r], and significance level [p] (* for < 5%, ** for <1%, *** for < 0.1%).

3 Results

The survey was conducted in February 2022. The collected and quality-controlled data set contains answers of 172 participants belonging to 123 different higher education institutions [He22b]. This is about a third of all research universities and universities of applied sciences in Germany. Please refer to [He22a] for a detailed description of data collection and quality assurance.

3.1 Judgements on perfect digitalization

As described above, participants were asked to rate the similarity between an imagined "perfect" digital institution and their own institution. We call this rating the Dx-score for simplicity. The Dx-score for the current state (Dx-now) and the expected future development within five years (Dx-5Y) were assessed and evaluated. To capture the judgements on an institutional level rather than an individual one, we averaged all ratings by participants of the same institution into one Dx-now and Dx-5Y score respectively. For about 70% of the institutions, only one person participated. Within the remaining 30% we captured data provided by 2-5 different people of the same institution. The score for those institutions therefore represents the combination of all captured perspectives of the CxO roles.

The overall Dx-score ratings increased considerably from the current state to the expected future state (on average, Dx-now=39.2 with standard deviation of 18.8 while Dx-5Y=66.5 with std deviation 16.2), as displayed in Figure 1.

The judgements of the current state of digital processes or the digital transformation of universities is often thought of as highly dependent on the level of competency and expectations as well as other factors. Only a few institutional factors prove to be a stable predictor for higher or lower Dx-scores. Performing a number of U-tests confirmed two

interdependences as displayed in Figure 1: on the left side between Dx-score and the age of the institution (n_1=57, n_2=56, m_1=30, m_2=40, β=49%, r=0,288, p=0,002**). For the Dx-score judgements predicting the state of the universities in five years from now, the U-test for the age of the institution also proved significant (n_1=58, n_2=58, m_1=65, m_2=74, β=39%, r=0,188, p=0,043*), as displayed on the right side of Figure 1. All other factors (size of the institution, type of institution, role of the participant, etc.) did not offer a consistent effect on the judgements of the Dx-score.

In addition, the estimates of costs invested for the digital transformation in the past three years (n_1=46, n_2=52, m_1=63, m_2=72, β=56%, r=0,237, p=0,019*) and for the next three years (n_1=44, n_2=54, m_1=65, m_2=70, β=47%, r=0,212, p=0,036*) revealed higher estimates for the Dx-score judgements as a positive correlation. However, both cost estimates also highly correlate with the complexity of the institution measured in terms of the Global Complexity Index (GCI) [Be16], which might be of interest as a future target of analysis.

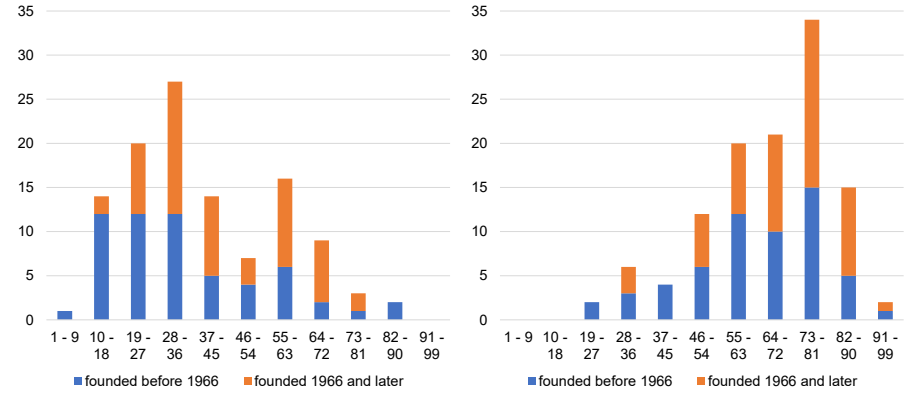


Figure 1: Histogram of number of institutions per Dx-score judgement today (left side) and in five years (right side).

We therefore conclude that the Dx-score judgements are potentially suitable for measuring the degree of digitalization without depending on other obvious factors of the universities such as size or form. It should be stressed that observing predictors confirmed by U-Tests does not suggest any causal relation between the two judgements. Further research is needed to determine the underlying functional relationships.

3.2 Interdependencies between Dx-scores and ESA factors

The interdependencies between the Dx-score ratings and the ESA factors were further evaluated. Here the data set was split into two groups at the median of the Dx-score judgements and the U-tests were performed against the various ESA factors. Table 1 in the appendix summarises all assessed factors and findings.

The reverse direction of having the ESA factors act as predictors of the Dx score was also calculated. Here the data set was split again at the median of the ESA factor, and the U-tests were performed against the two Dx-scores. Table 2 in the appendix summarizes all results.

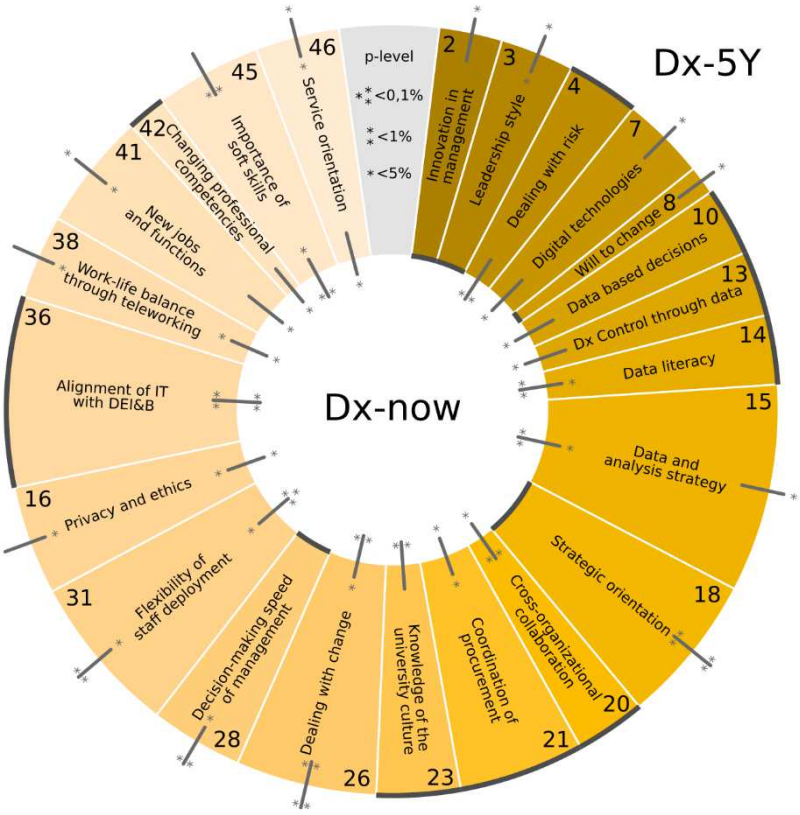


Fig. 2: Interdependencies between the ESA factors and Dx-now and Dx-5Y, respectively. The sizes of the segments correspond to the cumulated effect size of the statistically significant U-tests. Statistical significance is indicated using the * notation.

Both sets of results in terms of interdependencies are displayed in Figure 2 as an overview. All other assessed factors (see appendix) are far from being predictive relations and should be questioned in the context of understanding the nature of "digital transformation" in HE.

4 Discussion

The discussion of results is split into two sections. First, we explore Research Question 1 (RQ1) for the ESA factors which do not show any interdependencies with the Dx-scores.

Next, we focus on RQ2, first summarizing the observations systematically and then setting into a suitable context for a broader understanding of the nature of the terminology of "digital".

4.1 Why we do not observe interdependencies

Since both the Dx-score and the ESA-factors try to capture various aspects of the digital transformation of HE institutions, one should expect some interrelation. This might either be observed as a control variable to measure Dx (e.g., Dx-score as a predictor of one specific ESA factor) or as an independent variable which is related to any control aspect of the ongoing change (e.g., a specific ESA factor as a predictor of Dx-score). If ESA factors do not display any expected interdependencies, we need to carefully assess potential reasons for this observation.

Of all 36 assessed ESA factors, a quarter (#1, #11, #19, #24, #30, #33, #35, #37, #39) failed to show any interdependencies with either of the Dx-scores. An additional four factors showed only marginally significant results ($0.05 < p < 0.1$) in at least one of the performed U-tests (#6, #, 27, #29, #43).

The observed data could just be noisy due to participants misunderstanding either of the Dx-score, the respective ESA factor or both. Misunderstanding the Dx-score itself would lead to un-correlated results in most other factors as well, which we have not observed; ¾ of all ESA factors show some interdependency. If pure wording is the root cause, the Self-Assessment form should undergo careful revisions to enable the leadership of HE institutions to interact better with the statements and position themselves within the spectrum.

In addition, the amount of data collected might not have been sufficient to guarantee findings of medium or bigger effect sizes. Based on Cohen's d for the probability of 70%, finding medium effects requires about 50 samples; for larger effects, a sample size of only 20 is needed. Indeed, 9 and 22 ESA factors respectively would meet this criterion, among which eight factors do not show an interdependence. In other words, with the exception of ESA factor #39, we had a chance of about 60% to find medium size effects and above 90% to find large effects. It therefore seems unlikely that we have missed existing correlations for those eight factors based on the assumption of insufficient data samples.

Based further on the assumption of a high validity of the expertise indicated by the participants in the initial questions, we have to conclude the eight ESA factors to be of arguable use for the judgement of the digital transformation of higher education institutions in Germany. However, this is not in alignment with a major part of the literature on IT management [DS16], enterprise architecture [He17, GCC20], governance [Wi17, SB18, GBW21] and digital transformation [He17, He19]. As these factors are often discussed in the literature as having significant consequences for the digital transformation of institutions, we either have to search for other plausible reasons not to find the expected results, or question the often-cited literature based on our findings.

4.2 What does it mean to observe interrelations?

Focusing now on the results concerning RQ2, several observations can be made, both about the directional differences in the interdependencies between the ESA factors and the Dx-scores, and about the differences between the Dx-now and Dx-5Y interdependencies.

- The number of ESA factors which are predicted by the Dx-scores is higher than the number of ESA factors predicting the Dx-scores: 23>15. In other words, the Self-Assessment statements contain more factors modulated by the digital transformation than actively used to control it.

- All ESA factors which were predicted by one of the Dx-scores (in total 19 significant U-test results) were also predictors of the same Dx-score (twice on a marginal level of significance). In other words, we found a symmetrical interdependence of these 15 ESA factors for either Dx-now or Dx-5Y. This suggests people share a common understanding of these ESA factors and the perceived Dx-score, or both depend on a common root cause which might establish a correlation in the observations. These factors might also be seen as accelerators, if both interdependency directions are based on a causal relation. Reflecting on the model for scaling digital platforms [He19], they should be considered in future research and theoretical modelling.

- The alternative subset of 8 factors (=23-15) is correlated only in the direction towards the Dx-scores. These factors (#2, #4, #7, #8, #10, #13, #23, #42) might have the potential of controlling or modulating digital transformation, but would not be among the self-acceleration factors like the other 15 factors previously discussed. However, if digital transformation is considered to be more than observation, but could indeed be driven forward by management activities, these ESA factors might be of value without being influenced by the digital transformation itself.
  As ESA factors #2 and #8 are only interdependent to the future Dx-5Y, they might be of more crucial importance, since their results might not be visible immediately. Five of the ESA factors discussed here might be valuable for observing the current state of the digitally transformed institution, but might not show any future relevance.
  Only ESA factor #7, due to its nature as a long-term investment, will be relevant for both current and future digital transformation. It somehow is self-explanatory that digital transformation is indeed coupled to some degree to the use of digital technologies. However, this also supports the notion of measuring a relevant objective using the stated questions in the first place.

- The number of interdependencies observed for the Dx-now score is higher than the number of interdependencies observed for the Dx-5Y score (10>9 for Dx predicts ESA; 18>12 for ESA predicts Dx).
  This observation is consistent with the statistical effects of the Dx-5Y score appearing more focused; it contains less variability than the Dx-now score. Observing medium and large effects therefore is less likely for the Dx-5Y than for

the Dx-now. This might be a consequence of participants being consistently more optimistic about the future but less susceptible to real differences in the perceived situation. Alternatively, the current Dx-now score might appear more distributed because people differentiate more clearly due to a complex and realistic perception of the current situation.

- A subset of nine ESA factors is only correlated with the Dx-now score, but display no interdependency with the Dx-5Y ratings. Thus, these factors (#4, #10, #13, #14, #20, #21, #23, #36, #42) appear to be less relevant for changing the future Dx development.
  Again, the previously discussed difference in variance might be part of the reason as a pure statistical effect. However, for factors #14 and #20, this still seems unlikely due to the high sample size (>70). We would have expected, based on Cohen's d, less than a 1% chance of missing a large effect or a 20% chance of medium effect sizes, which we have observed for both for the reverse interdependency. The lack of observed interdependency therefore seems to constitute a true doubt of the participants that these factors will bear future effects. This suggests the need for further analysis, since it would be in contradiction of both the common belief in the relevance of data literacy and recent reports on the positive effects of cross-organizational collaboration.

- Another subset of five ESA factors (#2, #3, #8, #18, #28) is not interdependent with Dx-now but only shows correlations towards Dx-5Y ratings. In other words, these factors are less likely to be of use in judging the current Dx-state, but may be of relevance for the future development. The underlying wording of the statements may have pushed this asymmetry further.
  Changing these factors within the current situation is less likely to show quick wins or immediate effects in the digital transformation. In the long run, they are expected to be crucial for the digital transformation. As factors #2 and #8 again are not predicted by the future Dx-5Y score, they appear not to be directly part of the descriptive nature of being digital. Thus, they constitute independent factors of potential high relevance in changing the attitude towards Dx and the perception of the Dx without being mistaken for the Dx development itself.

- A third subset of five ESA factors (#15, #26, #31, #41, #46) displays interdependence towards both the Dx-now and the Dx-5Y, and at the same time is itself predicted by at least one or both Dx-score ratings. This indicates either a common understanding where these factors are just a part of understanding the nature of the Dx or, if the factors are seen as independent controllable entities, they might be again relevant for an increased, self-reinforcing acceleration of change. Working towards changing the current self-judgement within the universities should result in perceivable changes both now and in the long run.

Although the original Self-Assessment was structured into six topics, there appears to be no pattern in which any of those domains are statistically different form the others. This is consistent with Educause's view of a truly multi-factorial phenomenon. In [Re20] the

authors stress the relevance of change being distributed across all topics and domains. This also supports the requirement for a broad set of tools and cultural backgrounds to be orchestrated into one overall plan of change. A broad mindset seems to be required since not just a few, but many aspects appear to have profound relevance.

5 Summary

Understanding the nature of the digital transformation in higher education reveals a multi-factorial constellation. Educause's notion of a complex change in culture, technologies and the way we work has been challenged by our quantitative survey and subsequent statistical analysis. The proposed statements in Educause's Self-Assessment (ESA) for the Digital Transformation (Dx) were set in relation with an idealized digitalization score of today and an imaged future. Differences between the dimensions were assessed by correlations. Performing U-tests on data collected from one third of the German higher education ecosystem revealed major differences in how the various factors may contribute to the potential to observe or control the digital transformation.

On one hand, surprisingly, about 40% of the ESA factors (14 of the 37 evaluated) did not display any significant correlation. Half of these should have displayed existing medium or large effects, given a high probability for these effects to be found due to the participation rate and the sample size. In particular, the lack of interdependence of governance, user experience, and institutional differentiation is not consistent with other findings and common beliefs in IT management literature. Thus, this at least points towards questionable perspectives on the digital transformation by the participants of the evaluated survey.

On the other hand, this analysis revealed, or rather confirmed, Educause's expectation of the relevance of the other 23 topics in the observation and management of the digital change. As many of these factors display similar patterns, the analysis presented here is not suitable to point towards independent dimensions of the Dx journey. Further research and analysis is needed to clarify the grouping of the confirmed relevant factors into truly independent (and hopefully controllable or measurable) dimensions.

The relative sizes of the observed effects do point towards critical factors in the digital transformation. The top four factors (effect size above average of this analysis), which also having the biggest sample size, are the capability and attitudes concerning "dealing with change (#26)", "flexibility of staff deployment (#31)", "new jobs and functions (#41)", and "importance of soft skills (#45)". These factors are expected to gain further attention in the strategic discussions among IT leaders and CxOs in higher education.

6 Acknowledgements

The authors would like to thank all members of the group of experts for their active participation and support for the implementation and evaluation of the survey. Thanks are due to all participants for the time and effort they invested and for their willingness to assess the situation honestly and completely. Thanks also to the stakeholder associations for their support in sharing the survey with their members, as without them such a broad participation would hardly have been possible.

Bibliography

[Ab15] Abolhassan, F.: Was treibt die Digitalisierung? Warum an der Cloud kein Weg vorbeiführt, Springer, Wiesbaden, 2015. http://link.springer.com/content/pdf/10.1007/978-3-658-10640-9.pdf, last accessed 12.07.2022.

[AH22] Auth, G.; von der Heyde, M.: Die Rolle des Chief Digital Officers für die digitale Transformation von Hochschulen, HMD Praxis der Wirtschaftsinformatik, no. 345, Apr. 2022, doi: 10.1365/s40702-022-00869-6.

[Be16] Bergström, J. et al.: Putting the CHEITA Global Complexity Index to the test, in EUNIS 2016: Crossroads where the past meets the future, Thessaloniki, Greece, Juni 2016, Bd. 2016–1, S. 218–220 [Online]. See: https://www.eunis.org/eunis2016/wp-content/uploads/sites/8/2016/03/EUNIS2016_paper_49.pdf, last accessed 12.07.2022.

[DS16] Degkwitz, A.; Schirmbacher, P. (editors): Informationsinfrastrukturen im Wandel. Informationsmanagement an deutschen Universitäten. BOCK + HERCHEN Verlag, 2015. See: http://www.dini.de/fileadmin/docs/DINI_Informationsinfrastrukturen.pdf, last accessed 12.07.2022.

[GBW21] Gilch, H.; Book, A.; Wannemacher, K.: Kooperationen zur Digitalisierung in Lehre, Forschung und Verwaltung an den Hochschulen. Sekundärauswertung einer bundesweiten Erhebung“, in Digitalisierung in Studium und Lehre gemeinsam gestalten: Innovative Formate, Strategien und Netzwerke, Wiesbaden: Springer Fachmedien, 2021, S. 125–138. doi: 10.1007/978-3-658-32849-8_8.

[GCC20] Gomes, R.; da Cruz, A. M. R; Cruz, E. F.: EA in the Digital Transformation of Higher Education Institutions, in 2020 15th Iberian Conference on Information Systems and Technologies (CISTI), 2020, S. 1–6.

[Gr21] Graf-Schlattmann, M.; Meister, D. M.; Oevel, G.; Wilde, M.: Digitalisierungsstrategien auf dem Prüfstand. Eine empirische Untersuchung auf Basis der Grounded-Theory-Methodologie an deutschen Hochschulen, pedocs, 2019. doi: urn:nbn:de:0111-pedocs-180055.

[He17] von der Heyde, M; Auth, G.; Hartmann, A.; Erfurth, C.: Hochschulentwicklung im Kontext der Digitalisierung - Bestandsaufnahme, Perspektiven, Thesen, in INFORMATIK 2017, Chemnitz, Germany, Sep. 2017, Bd. 275, S. 1757–1772, doi: 10.18420/in2017_175.

[He19] von der Heyde, M; Auth, G.; Hartmann, A.; Erfurth, C.: Digital Transformation of Higher Education - Enabling Scaling Platforms, iCERi 2019, Seville, Spain, Nov. 2019, Bd. Proceedings of ICERI2019 Conference, S. 2107–2115, doi: 10.21125/iceri.2019.0584.

[He22a] von der Heyde, M.: Data collection and quality assurance of the survey on the digital transformation of business processes at universities in Germany. Zenodo, 2022, doi: 10.5281/zenodo.6383774.

[He22b] von der Heyde, M.: Data set of the survey on the digital transformation of business processes at universities in Germany. Zenodo, 2022, doi: 10.5281/zenodo.6383770.

[HG22] von der Heyde, M.; Gerl, A.: Entwicklungsstand der CIO-Funktion und hochschulübergreifenden IT-Governance im Kontext der Digitalen Transformation an Hochschulen in Bayern, HMD, no. 345, Apr. 2022, doi: 10.1365/s40702-022-00872-x.

[KHV+13] Keuper, F.; Hamidian, K.; Verwaayen, E.; Kalinowski, T.; Kraijo, C. (Hrsg.): Digitalisierung und Innovation – Planung – Entstehung Entwicklungsperspektiven, Springer Gabler, Wiesbaden, 2013.

[Mc21] McCormack, M.: EDUCAUSE QuickPoll Results: Institutional Engagement in Digital Transformation. EDUCAUSE Research Notes, See: https://er.educause.edu/articles/2021/8/educause-quickpoll-results-institutional-engagement-in-digital-transformation, last accessed 12.07.2022.

[Or20] Orr, D. et al.: A University Landscape for the Digital World, in Higher Education Landscape 2030: A Trend Analysis Based on the AHEAD International Horizon Scanning, D. Orr, M. Luebcke, J. P. Schmidt, M. Ebner, K. Wannemacher, M. Ebner, und D. Dohmen, Hrsg. Cham: Springer International Publishing, 2020, S. 1–4. doi: 10.1007/978-3-030-44897-4_1.

[Re20] Reinitz, B.: Consider the Three Ds When Talking about Digital Transformation, 2020. [Online]. Available at: https://er.educause.edu/blogs/2020/6/consider-the-three-ds-when-talking-about-digital-transformation, last accessed 12.07.2022.

[SB18] Schünemann, I.; Budde, J.: Hochschulstrategien für die Lehre im digitalen Zeitalter - Keine Strategie wie jede andere!, Hochschulforum Digitalisierung (HFD), Arbeitspapier 38, Sep. 2018 [Online]. See: https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr38%20Empfehlungen_Strategieentwicklung_WEB.pdf, last accessed 12.07.2022.

[UA17] Urbach, N.; Ahlemann; F.: Die IT-Organisation im Wandel: Implikationen der Digitalisierung für das IT-Management. HMD Praxis der Wirtschaftsinformatik, 2017, doi: 10.1365/s40702-017-0313-6.

[We15] Weinman, J.: Digital disciplines. Attaining market leadership via the cloud, big data, social, mobile, and the internet of things, Wiley (Wiley CIO series), Hoboken, 2015.

[Wi17] Wimmer, M.: IT-Governance an Hochschulen. Notwendigkeit, Stand und Wege zum Erfolg, Die Hochschule: Journal für Wissenschaft und Bildung, Bd. 26, Nr. 01/2017, S. 70–82, Jan. 2017, doi: urn:nbn:de:0111-pedocs-166375.

Appendix

#	Topic	Dx-score rating of today predicts ESA								Dx-score rating in 5 years today predicts ESA							
		n1	n2	median 1	median 2	power	effect size	diff-U	signif. level	n1	n2	median 1	median 2	power	effect size	diff-U	signif. level
	Strategic Innovation																
1	Institutional differentiation	34	36	63,9	66,8	0,121	0,035	-142	0,769	32	40	64,0	69,2	0,231	0,116	-86	0,324
2	Innovation in management	9	10	37,7	62,5	0,965	0,431	-2	0,060	8	11	42,8	61,8	0,808	0,398	-3	0,083
3	Leadership style	22	22	66,8	69,0	0,105	0,050	-70	0,742	22	23	63,6	75,1	0,528	**0,305**	3	**0,041**
4	Dealing with risk	10	7	39,3	51,6	0,444	0,213	-12	0,380	10	7	46,8	49,3	0,099	0,024	-20	0,922
6	Strategic contribution of IT	9	10	52,4	62,5	0,362	0,150	-17	0,514	9	10	50,8	72,9	0,883	0,393	-4	0,086
7	Digital technologies	20	22	33,9	44,8	0,502	0,260	-11	0,092	22	20	35,0	46,5	0,529	0,272	-8	0,078
8	Will to change	40	37	56,6	64,2	0,325	0,112	-97	0,328	38	38	56,2	66,7	0,465	0,199	-22	0,083
	Data and Analytics																
10	Data based decisions	5	9	36,0	48,7	0,535	0,267	-8	0,317	7	7	35,3	53,0	0,748	0,427	-3	0,110
11	Maturity of data and interfaces	18	20	42,6	48,2	0,276	0,171	-32	0,293	18	21	45,9	45,9	0,003	0,027	-64	0,866
13	Dx Control through data	9	11	23,3	38,8	0,708	0,348	-6	0,119	10	10	24,8	41,6	0,813	0,389	-3	0,082
14	Data literacy	37	34	29,6	45,8	0,610	**0,287**	39	**0,016**	34	36	32,6	43,5	0,415	0,212	-16	0,076
15	Data and analysis strategy	8	10	17,9	39,4	1,002	**0,545**	3	**0,021**	9	8	22,2	42,0	0,948	0,397	-4	0,102
	Institutional Alignment																
18	strategic orientation	6	10	44,5	54,7	0,419	0,271	-9	0,278	9	8	38,3	67,1	1,206	**0,630**	6	**0,009**
19	impact of governance	19	24	52,9	62,0	0,374	0,157	-39	0,304	20	22	54,0	61,8	0,316	0,120	-47	0,435
20	cross-organizational collaboration	36	37	53,1	69,9	0,815	**0,365**	105	**0,002**	35	40	60,0	64,4	0,223	0,135	-75	0,243
21	coordination of procurement	7	9	21,3	54,2	1,269	**0,542**	1	**0,030**	9	8	33,0	51,2	0,696	0,280	-9	0,248
23	knowledge of the university culture	8	5	58,8	75,4	0,748	0,365	-5	0,188	4	8	72,8	66,0	0,336	0,049	-11	0,865
24	Adjustment of the financing of IT	21	25	41,5	43,3	0,068	0,047	-75	0,749	22	24	44,8	41,5	0,130	0,055	-73	0,709
	Flexibility and Agility																
26	Dealing with change	39	32	42,6	55,6	0,559	**0,281**	35	**0,018**	35	36	42,8	56,2	0,591	**0,315**	60	**0,008**
27	Flexibility and agility	8	10	26,9	48,1	0,948	0,461	-1	0,051	8	10	33,4	49,3	0,754	0,398	-4	0,091
28	Decision-making speed of management	20	25	52,3	64,2	0,500	0,266	-8	0,075	21	24	52,7	69,2	0,765	**0,363**	20	**0,015**
29	Enterprise Architecture	10	10	14,9	28,0	0,718	0,338	-6	0,131	10	9	19,9	25,4	0,305	0,131	-18	0,568
30	Procurement of IT	26	22	68,6	74,9	0,248	0,060	-75	0,679	27	21	69,1	72,8	0,149	0,061	-74	0,670
n.a	Operation of IT	10	9	65,7	68,0	0,101	0,094	-20	0,683	10	8	60,8	70,3	0,428	0,230	-12	0,328
31	Flexibility of staff deployment	32	28	46,5	56,5	0,470	**0,295**	21	**0,022**	30	30	45,5	57,6	0,562	**0,284**	16	**0,028**
	Diversity, Equity, Inclusion and Belonging																
16	Privacy and ethics	25	22	49,7	63,2	0,530	**0,289**	1	**0,047**	26	22	53,6	62,7	0,349	0,266	-6	0,066
33	User orientation	20	21	51,5	60,6	0,468	0,143	-41	0,361	20	21	58,1	58,8	0,036	0,012	-73	0,938
35	Diversity, equity and participation	18	23	61,5	63,6	0,101	0,070	-58	0,655	23	19	63,1	63,3	0,012	0,029	-71	0,850
36	Alignment of IT with DEI&B	6	6	26,8	75,7	1,627	**0,832**	5	**0,004**	3	9	45,3	63,1	-	-	-	-
37	Flexible staff development	23	22	43,5	48,7	0,238	0,091	-60	0,540	23	20	43,5	51,3	0,368	0,171	-35	0,263
38	Work-life balance through teleworking	37	36	67,6	75,5	0,452	**0,235**	4	**0,045**	37	38	68,1	74,7	0,380	**0,245**	15	**0,034**
39	Transfer of DEI knowledge	8	9	61,0	61,2	0,008	0,000	-21	1,000	9	9	59,1	65,2	0,287	0,219	-12	0,354
	Transformation of Work and Skills																
41	New jobs and functions	19	22	44,2	55,1	0,505	0,261	-11	0,094	22	21	46,3	57,3	0,486	**0,300**	0	**0,049**
42	Changing professional competencies	22	21	39,5	48,1	0,522	0,282	-5	0,065	21	20	43,8	45,7	0,119	0,057	-62	0,715
43	Facilitation functions of IT personnel	10	9	43,8	47,3	0,163	0,037	-23	0,870	9	9	45,3	50,6	0,270	0,177	-14	0,453
45	Importance of soft skills	39	32	48,5	61,9	0,653	**0,261**	20	**0,028**	35	35	49,5	62,0	0,632	**0,352**	83	**0,003**
46	Service orientation	31	37	62,2	68,9	0,441	0,220	-12	0,069	32	38	62,4	68,8	0,432	**0,235**	0	**0,049**

Table 1: Dx-scores as predictors of Educause self-assessment factors

Table 2: Educause self-assessment factors as predictors of Dx-scores

#	Topic	ESA predicts Dx-score of today								ESA predicts Dx-score rating in five years							
		n1	n2	median 1	median 2	power	effect size	diff-U	signif. level	n1	n2	median 1	median 2	power	effect size	diff-U	signif. level
Strategic Innovation																	
1	Institutional differentiation	36	34	35,0	33,7	0,029	0,027	-148	0,823	35	37	70,0	72,0	0,221	0,104	-96	0,376
2	Innovation in management	9	10	27,0	52,5	0,894	0,431	-2	0,060	8	11	58,0	82,0	1,087	**0,549**	5	**0,017**
3	Leadership style	23	21	30,0	35,0	0,067	0,051	-69	0,733	22	23	62,5	75,0	0,619	**0,338**	13	**0,023**
4	Dealing with risk	9	8	25,0	50,8	1,222	**0,630**	6	**0,009**	8	9	50,5	67,0	0,810	0,350	-6	0,149
6	Strategic contribution of IT	10	9	31,5	46,5	0,299	0,169	-16	0,462	9	10	50,0	76,0	0,941	0,412	-3	0,072
7	Digital technologies	21	21	26,5	44,0	0,600	**0,355**	13	**0,021**	20	22	57,0	72,0	0,829	**0,381**	20	**0,014**
8	Will to change	38	39	34,0	33,0	0,110	0,053	-147	0,639	36	40	60,0	72,0	0,565	**0,249**	20	**0,030**
Data and Analytics																	
10	Data based decisions	7	7	27,0	61,0	1,175	**0,598**	2	**0,025**	7	7	50,0	75,0	0,946	0,427	-3	0,110
11	Maturity of data and interfaces	14	24	30,0	35,5	0,505	0,226	-19	0,164	15	24	75,0	68,5	0,059	0,032	-61	0,840
13	Dx Control through data	9	11	26,5	55,0	1,043	**0,569**	7	**0,011**	9	11	51,0	75,0	0,737	0,280	-10	0,210
14	Data literacy	34	37	26,0	40,0	0,579	**0,318**	62	**0,007**	32	38	62,5	71,5	0,445	0,189	-33	0,114
15	Data and analysis strategy	9	9	25,0	55,0	1,549	**0,760**	14	**0,001**	8	9	55,5	77,0	1,012	**0,513**	1	**0,034**
Institutional Alignment																	
18	strategic orientation	8	8	33,0	52,5	0,864	0,420	-3	0,093	8	9	50,5	77,0	1,412	**0,677**	8	**0,005**
19	impact of governance	22	21	32,5	40,0	0,247	0,156	-39	0,308	21	21	67,0	72,0	0,399	0,146	-41	0,346
20	cross-organizational collaboration	34	39	30,0	40,0	0,486	**0,247**	13	**0,035**	34	41	68,5	71,0	0,257	0,136	-74	0,237
21	coordination of procurement	8	8	28,5	58,0	1,197	**0,551**	2	**0,027**	8	9	49,8	77,5	0,831	0,420	-3	0,083
23	knowledge of the university culture	5	8	20,8	48,3	1,353	**0,731**	4	**0,008**	4	8	63,0	76,0	0,750	0,294	-6	0,308
24	Adjustment of the financing of IT	22	24	35,5	35,2	0,034	0,006	-88	0,965	22	24	71,0	69,5	0,049	0,010	-87	0,947
Flexibility and Agility																	
26	Dealing with change	34	37	28,2	40,0	0,710	**0,316**	60	**0,008**	33	38	60,0	72,5	0,717	**0,316**	60	**0,008**
27	Flexibility and agility	8	10	28,5	52,5	0,913	0,461	-1	0,051	7	11	50,0	77,0	0,920	0,438	-2	0,063
28	Decision-making speed of management	23	22	30,0	42,0	0,372	0,254	-12	0,089	21	24	60,0	75,0	0,905	**0,448**	45	**0,003**
29	Enterprise Architecture	10	10	26,8	48,3	0,687	0,389	-3	0,082	9	10	66,0	68,5	0,272	0,037	-23	0,870
30	Procurement of IT	24	24	30,0	31,5	0,088	0,012	-92	0,934	25	23	65,0	66,0	0,192	0,100	-62	0,489
n.a.	Operation of IT	9	10	30,0	52,5	0,926	0,375	-5	0,102	9	9	60,0	75,0	0,887	0,427	-2	0,070
31	Flexibility of staff deployment	29	31	25,0	44,0	0,791	**0,455**	106	**0,000**	28	32	57,5	71,5	0,815	**0,365**	58	**0,005**
Diversity, Equity, Inclusion and Belonging																	
16	Privacy and ethics	22	25	27,5	40,0	0,637	**0,339**	17	**0,020**	21	27	60,0	72,0	0,623	**0,322**	13	**0,025**
33	User orientation	19	22	30,0	37,5	0,454	0,249	-14	0,111	18	23	69,5	70,0	0,321	0,172	-33	0,270
35	Diversity, equity and participation	20	21	35,7	35,0	0,169	0,053	-63	0,735	20	22	68,0	65,5	0,120	0,062	-62	0,687
36	Alignment of IT with DEI&B	6	6	25,8	52,5	1,624	**0,832**	5	**0,004**	5	7	70,0	82,0	0,544	0,445	-3	0,123
37	Flexible staff development	22	23	28,5	35,0	0,165	0,129	-49	0,388	21	22	60,0	69,5	0,359	0,145	-42	0,343
38	Work-life balance through teleworking	34	39	29,7	40,0	0,515	**0,252**	17	**0,031**	36	39	60,0	71,0	0,387	0,208	-15	0,071
39	Transfer of DEI knowledge	9	8	36,0	39,8	0,157	0,093	-17	0,700	9	9	65,0	77,0	0,753	0,343	-6	0,145
Transformation of Work and Skills																	
41	New jobs and functions	22	19	29,0	44,0	0,683	**0,343**	9	**0,028**	21	22	60,0	72,0	0,797	**0,374**	20	**0,014**
42	Changing professional competencies	21	22	27,0	37,7	0,598	**0,341**	11	**0,025**	20	21	62,5	70,0	0,385	0,155	-38	0,322
43	Facilitation functions of IT personnel	9	10	26,5	48,3	0,685	0,375	-5	0,102	8	10	50,5	73,5	0,937	0,398	-4	0,091
45	Importance of soft skills	35	36	25,0	37,7	0,594	**0,314**	59	**0,008**	33	37	60,0	72,0	0,475	0,233	-2	0,052
46	Service orientation	34	34	30,0	44,0	0,513	**0,242**	3	**0,046**	34	36	66,0	75,0	0,566	**0,256**	15	**0,032**

Bildungsinformatik als Accelerator innovativer Veränderungsprozesse in der Hochschule der Zukunft

Über die Notwendigkeit und Potenziale eines interdisziplinären Studien- und Forschungsfelds

Kevin Saukel[1]

Abstract: Im Frühling wurden die Hochschulen zu Reallaboren digitalen Lehren und Lernens. Nach nun fast zwei Jahren beständigem Emergency Teachings nimmt nun die Vision einer Blended University mehr und mehr an Gestalt an, was den Anlass dafür gibt die Verzahnung von Bildungstechnologien mit Studium und Lehre unter dem Blickwinkel von Hochschul- und Forschungsstrukturen zu betrachten. Mit dem Blick auf weiter stattfindende Veränderungsprozesse in der Digitalität der Hochschulen und der Einführung von hybriden Lernstrukturen sind ferner diese neuen Strukturen sowohl aus informationswissenschaftlicher und hochschuldidaktischer Perspektive zu betrachten. Hieraus bilden sich komplexe Herausforderungen, die einer stärkeren wissenschaftlichen Betrachtung und Forschung bedürfen. Genau hier soll die Bildungsinformatik ansetzen und neue Strukturen zur Auseinandersetzung mit diesen Querschnittsthemen schaffen.

Keywords: Bildungsinformatik; Studiengang; Educational Technologies; Forschung;

1 Einleitung

Praktisch über Nacht wurden die Hochschulen im Frühling 2020 zu realen Experimentierfeldern der Hochschullehre. Mittlerweile hat sich nach 2 Jahren des Emergency Teachings bei vielen Akteuren eine Gewöhnung an digitaler Hochschullehre abgezeichnet, die jedoch nach Beginn des Sommersemester 2022 wieder abnimmt und Gefahr läuft in Vergessenheit zu geraten [HO22]. Dass viele Hochschulen nun öffentlich das Sommersemester 2022 als „Präsenzsemester“ bezeichnen, lässt vermuten, dass die post-pandemische Vorstellung von Studium und Lehre sich in der Dualität von Präsenz und Distanz widerspiegelt und sich der Wunsch nach einer Präsenzuniversität durchsetzt. Demgegenüber stellt sich eine Befragung des CHE Masterrankings 2021 und des CHE Hochschulrankings 2021, in welchem nur ein kleiner Anteil von Studierenden und Professor*innen sich für die reine Präsenzlehre aussprechen [Ho22:19]. Vielmehr äußert sich der Wunsch darin künftig digitale Elemente in die Präsenzlehre einzubetten und zu großen Teilen auch Blended-Learning und Hybride-Formate in Studium und Lehre zu

[1] Johann Wolfgang Goethe-Universität, Technische Hochschule Mittelhessen, ehrenamt@saukel-kevin.de

doi:10.18420/inf2022_119

implementieren. Aus Perspektive von Studierenden und Professor*innen liegt die Zukunft der Lehre daher in der Mischung von Präsenz- und Distanzelementen.

Die vergangenen zwei Jahre zeigten uns, dass wir dazu in der Lage sind mit neuen Lernformaten zu experimentieren. Die Ad-hoc implementierten Lösungen in der Hochschullehre speisten sich jedoch meist aus einfachen technischen Lösungen und ähnelten überwiegend den Präsenzformaten [BFS22:14]. Dies lässt den Schluss zu, dass die didaktische Nutzung von Lern- und Bildungsanwendungen in der Zukunft von Studium und Lehre noch ausbaufähig und das Potenzial noch nicht ausgeschöpft ist.

Aus dieser Grundlage heraus ergeben sich für die zukünftige Implementierung technischer Lösungen in der Hochschullehre weitreichende und komplexe Anforderungen, die in der Zukunft interdisziplinär adressiert werden müssen. Hierzu soll in diesem Beitrag auf die Einführung eines Studienganges Bildungsinformatik und dessen Rolle in Forschung und Lehre eingegangen werden.

Das Themenfeld Bildungsinformatik kann dabei hermeneutisch im Querschnitt der informations- und bildungswissenschaftlichen Disziplinen verortet werden und adressiert damit einerseits den Einsatz und die Bewertung von Technologien und andererseits dessen Entwicklung ausgerichtet an bildungswissenschaftlichen Erkenntnissen. Am Besten lässt sich dies durch den stärker werdenden Einzug von „educational technologies“ [STI21] in Bildungskontexten (z.B. Gamification, LMS, VR/AR based Learning, Learning Analytics) darstellen. Insbesondere hier besteht die Herausforderung und Forschungsfrage, wie solche educational technologies entwickelt und genutzt werden müssten, um Lernprozesse optimal fördern und an den Lernenden ausrichten zu können.

2 Rolle der Bildungsinformatik

Mit dem Entwurf der EdTech-Charta vom Stifterverband werden aktuelle Empfehlungen im Bereich der Bildungsinnovationen diskutiert [STI21], die darauf abzielen Rahmenbedingungen für Bildungsinnovationen in der Hochschule zu schaffen. Der Grundsatz besteht darin, dass die Arbeit von EdTechs stärker mit der Hochschule vernetzt und verzahnt werden muss, um von gegenseitigen Synergien zu profitieren und Bildungsinnovationen in Hochschule weiter voranzubringen. Laut dem Stifterverband ist dabei unter dem Begriff EdTechs folgendes zu verstehen:

> „Unter dem Begriff EdTech lassen sich innovative und technologieorientierte Unternehmen und Start-ups zusammenfassen, die Lösungen, Services und Produkte im Bereich der Lern- und Bildungsanwendungen anbieten.“ – [STI21]

Dabei kann ebenfalls die Annahme aufgestellt werden, dass in Zukunft (mit den weiter anhaltenden Digitalisierungsprozessen im Bildungsbereich) weitere Berufsfelder im Querschnitt von Bildung und Informatik entstehen. Nicht nur in den Hochschulen, sondern auch in Bereichen der Weiterbildung, schulischen Bildung und der Erwachsenenbildung

wird u.a. durch den Einsatz bereits vorhandener Lernplattformen und -tools erkenntlich, dass sowohl digitale Bildungsangebote und -formate als auch der Einsatz von educational technologies in der Zukunft zunehmen werden. Bereits jetzt sprechen wir (Stand: 2021) in Deutschland über ein Umsatzvolumen von ca. 100 Mio. € in der e-Learning Branche [STA21], wo diverse berufliche Tätigkeitsprofile wiederzufinden sind [MBB21:7], die im genannten Querschnitt agieren. Aktuell sind demgegenüber nur punktuelle Angebote und Arbeitsbereiche der Bildungsinformatik auffindbar. Aus ökonomischer und wissenschaftlicher Perspektive ist daher die Überlegung zur Einführung von Bildungsinformatik als Studiengang und Forschungsfeld konkreter in den Blick zu fassen.

2.1 Konzeption des Studienganges Bildungsinformatik

Die Grundlage eines Studiengangs Bildungsinformatik ist die Behandlung von Fragestellungen und Anwendungsfeldern im Zusammenspiel zwischen Bildungswissenschaften und Informatik. Insbesondere mit dem Einzug der Future Skills in den Hochschuldiskurs [STI22:50ff] müssen künftige Entwicklungen und Implementierungen von Bildungstechnologien (LMS, Lernapps und -tools, Governance Strukturen) unter dem kritischen Blickwinkel kompetenzorientierter Bildung betrachtet werden. Hierzu greift der Studiengang einerseits auf (evidenzbasierte) pädagogische und andererseits auf informationswissenschaftliche Forschung zurück. Ziel dieses Studienganges ist die Vermittlung von Fähigkeiten zum Entwickeln, Gestalten, Einsetzen und Bewerten von Informationssystemen auf Basis von softwaretechnischen, pädagogischen und (medien)didaktischen Gesichtspunkten und Kompetenzen.

Andreas Zendler (Prof. an der PH Ludwigsburg) hatte bereits 2005 in einem Artikel über die Konzeption von Bildungsinformatik im Rahmen eines online-gestützten interdisziplinären Studienganges folgende mögliche Module festgehalten [Ze05:5]:

- Bildungsinformatische Grundlagen
- Informatik
- Bildungswissenschaften
- Lehr-/Lernwissenschaften
- Methoden und Verfahren der Bildungsinformatik

Die von Zendler genannten Modulinhalte [Ze05:5ff] können als Inspiration einer weiteren Konzeption des Studienganges dienen. Sie müssen jedoch auch aus Erkenntnissen im technologischen und pädagogischen Bereich des letzten Jahrzehnts auf den Prüfstand gestellt und weiterentwickelt werden.

2.2 Bildungsinformatik als Accelerator von Bildungsinnovationen

Die Bildungsinformatik hat die Chance Herausforderungen aktueller und künftiger Bildungsinnovationen aus informatorischem und bildungswissenschaftlichem

Blickwinkel zu adressieren. Bereits eine kleine Auflistung von Themen der Bildungsinformatik gibt Aufschluss über die Diversität der möglichen Themenbereiche:

- Lernszenarios mithilfe von Virtual und Augmented Reality
- Learning Analytics
- Lernarchitekturen (hybrid, digital)
- Virtuelle Lernumgebungen (LMS, Lerntool und -apps)
- Lehren und Lernen in Hybriden und Blended-Learning Formaten
- Digitale Verwaltungs- und Governancestrukturen in Bildungsbereichen

Im Kontext weiterer Veränderungsprozesse innerhalb der Gesellschaft, Wissenschaft und Wirtschaft muss die Wissenschaft und Forschung auf künftige Ansprüche entsprechend reagieren können, weshalb ein Ausbau von Forschungsaktivitäten in diesen Querschnittsbereichen unerlässlich wird. Die Bildungsinformatik kann hierbei ein solcher Ort sein, in denen die Herausforderungen zur Implementierung und Anwendung von Bildungstechnologien im Bildungsbereich interdisziplinär bearbeitet werden können.

3 Die Zeit zu handeln ist jetzt

Aus den aktuellen Ansprüchen und Herausforderungen der Hochschulentwicklung stellt sich die Frage, wie künftige Entwicklungen im Bereich der Bildungsinnovationen fachlich und interdisziplinär begleitet und umgesetzt werden können. Hier setzt die Bildungsinformatik an und schafft einen strukturierten Raum für diese Themen. Der Diskurs um die Nutzung von educational technologies im Bildungsbereich kann hiermit auf forschungstheoretische Beine gestellt werden, die insbesondere für zukünftige Entwicklungen in diesem Bereich notwendig sein werden.

Der vorliegende Kurzbeitrag hat die Notwendigkeit und Sinnhaftigkeit der Einführung eines Studiengangs Bildungsinformatik und den Ausbau dazugehöriger Forschungsaktivitäten auf diesem Feld versucht zu erörtern. Zur Verbesserung der Daten- und Diskussionsgrundlage ist es daher im nächsten Schritt notwendig erstens Studiengänge und Forschungsfelder der Bildungsinformatik zu erfassen und zweitens integrativ in bereits vorhandene Disziplinen (z.B. Erziehungswissenschaft, Informatik) hineinzublicken, um feststellen zu können, inwieweit Themen der Bildungsinformatik dort bereits verortet sind.

Abschließend bleibt festzuhalten, dass wir durch die zweijährige Pandemie viele Erfahrungen gemacht haben, jedoch uns das auch die Erkenntnis gegeben hat, dass das Potenzial in Studium und Lehre (in zukünftigen Kontexten: Blended University, AR/VR, Hybrid Learning, Digitale Elemente in Studium und Lehre) in dieser Zeit noch nicht ausgeschöpft wurde. Für die zukünftige Hochschulentwicklung bedeutet dies, dass wir künftig viel stärker und tiefer diese Themenbereiche adressieren müssen, damit wir den Herausforderungen für die Hochschule der Zukunft gewachsen sind.

Literaturverzeichnis

[BFS22] Budde, Jannica; Friedrich, Julius-David; Sames, Josephine: Unsere Vision vom Blended Learning zur „Blended University“. In (Hochschulforum Digitalisierung, Hrsg.): strategie digital. Magazin für Hochschulstrategien im digitalen Zeitalter (03/2022), Ausgabe #02, Hochschulforum Digitalisierung, S. 13–20, 2022.

[HO22] Horndasch, Sebastian, Hochschulforum Digitalisierung: https://hochschulforumdigitalisierung.de/de/blog/keine-rolle-rueckwaerts-bitte, Stand: 15.05.2022

[Ho22] Horstmann, Nina: CHECK – Studienbedingungen an deutschen Hochschulen während der Corona-Pandemie. CHE gemeinnütziges Centrum für Hochschulentwicklung, Gütersloh, S. 19, 2022.

[MBB21] mbb Institut GmbH: Digitales Lernen floriert in der Pandemie. Ergebnisse des mmb-Branchenmonitors „E-Learning-Wirtschaft“ 2021. mmb Institut GmbH, Essen, S. 7, 2021.

[STA21] STA, Statista GmbH., https://de.statista.com/themen/1371/e-learning/, Stand: 15.05.2022.

[STI21] STI, Stifterverband für die Deutsche Wissenschaft e.V., https://www.stifterverband.org/edtech-charta, Stand: 15.05.2022.

[STI22] Stifterverband für die Deutsche Wissenschaft e.V.: Hochschul-Bildungs-Report 2020 – Abschlussbericht 2022. Stifterverband für die Deutsche Wissenschaft e.V., Essen, S. 50-58, 2022.

[Za05] Zendler, Andreas: Bildungsinformatik: Konzeption eines online-gestützten interdisziplinären Studienganges, S. 5-7, 2005.

„Und hier geht es übrigens jetzt gerade manuell weiter!“

Ergebnisse einer Expertenbefragung zur Digitalisierung in der Hochschulverwaltung

Stefanie Lemcke [1], Ulrike Lucke [2]

Abstract: Die Digitalisierung der Hochschulverwaltung ist Gegenstand zahlreicher strategischer Initiativen und wissenschaftlicher Studien. Während bisherige Arbeiten sich vorwiegend auf den unmittelbaren Prozess der Digitalisierung von Verwaltungsabläufen richten, stehen in der hier vorgestellten Studie die vorgelagerten Entscheidungsfindungsprozesse im Fokus. Im Rahmen einer qualitativen Untersuchung wurden IT-Entscheider aus der deutschen Hochschullandschaft befragt, wie die Auswahl der zu digitalisierenden Verwaltungsprozesse erfolgte und welche Einflussfaktoren bestanden. So wurde das wirkende Kriteriengeflecht offengelegt und in ein Kategoriensystem überführt, das für künftige Entscheidungsprozesse Unterstützung und Transparenz bietet.

Keywords: Digitalisierung, Hochschule, Verwaltungsprozesse, Entscheidungsprozesse, Kriterien.

1 Digitalisierung von Verwaltungsprozessen an Hochschulen

In den vergangenen zwei Jahrzehnten hat die Digitalisierung Einzug in die öffentliche Verwaltung gehalten und dort teils große Veränderungen ausgelöst [Du05, vdH18]. Die damit verbundenen Entscheidungs- und Steuerungsprozesse beinhalten eine Abkehr von den Ideen des New Public Management hin zur Reintegration von Organisationen, bedarfsorientierter Ganzheitlichkeit und kollaborativer Governance [MD13, Lu20]. Während sich die Aufmerksamkeit auf die Digitalisierung von Lehre und Studium [Lo20, Ha21] und inzwischen auch von Forschung [RfII21] richtet, bleibt die Digitalisierung der Hochschulverwaltung noch hinter den Erwartungen zurück [Le18].

Dabei ist eine wichtige Frage nicht nur, *wie* die digitale Transformation hier umzusetzen ist, sondern schon *wo* damit begonnen werden sollte. Vorgehensmodelle oder Entscheidungshilfen, die Orientierung in diesem komplexen Veränderungsprozess bieten könnten, fehlen noch. Sollen zuerst kleine oder große, einfache oder komplizierte, gut verstandene oder noch schwergängige Prozesse digitalisiert werden? Der Rückgriff auf Prozessmetriken mag als geeignetes Mittel erscheinen [Le21a], erweist sich aber wegen

[1] Universität Potsdam, CIO-Bereich, Am Neuen Palais 10, 14469 Potsdam, stefanie.lemcke@uni-potsdam.de, http://orcid.org/0000-0003-0043-6587
[2] Universität Potsdam, Institut für Informatik und Computational Science, An der Bahn 2, 14476 Potsdam, ulrike.lucke@uni-potsdam.de, http://orcid.org/0000-0003-4049-8088

doi:10.18420/inf2022_120

deren engem Fokus und unbekannten Schwellwerten als unzureichend, um einen Prozess auf seine Digitalisierbarkeit zu untersuchen. Zur Beantwortung der Frage, wann ein bzw. welcher Prozess digitalisiert werden sollte, ziehen Entscheider in der Praxis ein vielfältiges Kriteriengeflecht heran, das jedoch von außen nicht immer nachvollziehbar ist. Dieses Terrain soll deshalb mit einer Interview-Studie kartographiert werden, um die bestehenden Entscheidungsspielräume und -prozesse besser zu verstehen und damit inhärentes Wissen für künftige Entscheidungsprozesse nutzbar zu machen.

Das Ziel dieses Beitrags ist es, das theoretische Verständnis für derartige Entscheidungsprozesse zu vertiefen und die Wechselbeziehung zwischen digitaler Transformation und Governance-Fragen empirisch zu untersuchen. Es werden spezifische Herausforderungen, Strukturen und Dynamiken im Kontext der digitalen Transformation in der Hochschulverwaltung beleuchtet. Kapitel 2 beschreibt dafür die methodische Herangehensweise in Form einer Expertenbefragung und ihrer Auswertung. Die Ergebnisse dieser Analyse werden in Kapitel 3 ausschnittweise vorgestellt. Kapitel 4 fasst die Erkenntnisse zusammen und gibt einen Ausblick auf weitere Arbeiten.

2 Befragung von Digitalisierungsexperten

In einer Expertenbefragung sollte die Forschungsfrage beantwortet werden, wie an Hochschulen die Entscheidung gefällt wird, welche Verwaltungsprozesse digitalisiert werden. Das beinhaltete drei Unterfragen: welche Prozesseigenschaften die Digitalisierung begünstigen oder hemmen; welche Einflussfaktoren während der Digitalisierung existieren und welche Auswirkungen durch die Digitalisierung der Prozesse entstehen oder zumindest erwartet werden.

Den methodischen Rahmen der Interview-Studie bilden Verfahren aus der qualitativen Sozialforschung. Die Erhebung der Daten wurde mit Hilfe eines durch einen offenen Leitfaden strukturierten Experteninterviews durchgeführt. Experten sind Personen, die über ein Sonderwissen in Bezug auf bestimmte Wissensgebiete, z.B. über Zusammenhänge, Abläufe und Mechanismen innerhalb einer Organisation, verfügen oder es von anderen Personen zugeschrieben bekommen [PW14]. Die Strukturierung durch einen offenen Leitfaden ermöglicht dabei eine Eingrenzung auf bestimmte Forschungsfragen und bietet zugleich eine Flexibilität in Bezug auf den Verlauf des Interviews [PW14]. Insgesamt gliederte sich das Interview in vier Themenbereiche:

- In der *Selbstvorstellung* konnten die interviewten Personen ihre Rolle im Digitalisierungsvorhaben ihrer Einrichtung darlegen und einordnen.
- Im Block *Prozesse* wurden die Auswahl der zu digitalisierenden bzw. bereits digitalisierten Prozesse betrachtet sowie die Gründe und Einflussfaktoren (wie Prozesseigenschaften oder Rahmenbedingungen) hinterfragt.

- Der Bereich *Planung und Umsetzung* beleuchtete die Durchführung der Digitalisierung selbst, wobei auch hier Einflussfaktoren (wie Eigenschaften der Prozesse, Voraussetzungen, Rahmenbedingung und Entscheidungen) von Interesse waren.
- Abschließend wurden nach persönlichen *Erfahrungen* und damit verbunden in einer retrospektiven Betrachtung der Entwicklungen nach zusätzlichen Informationen gefragt, die in den sachlichen Schilderungen des vorangegangenen Blocks ausgelassen oder übersehen wurden.

Von acht angefragten haben sieben IT-Entscheider an der qualitativen Studie teilgenommen. Dies waren vier (ehemalige) CIOs, zwei leitende Personen von Verwaltungseinrichtungen und eine IT-Leitung. Sie waren tätig an zwei Universitäten, einer Fachhochschule, einer dualen Hochschule und zwei außeruniversitären Forschungseinrichtungen. Insgesamt waren Personen aus vier Bundesländern vertreten. Die Interviews wurden Anfang 2020 durchgeführt.

Die Auswertung der Interviews erfolgte mit Hilfe der qualitativen Inhaltsanalyse. Speziell wurde hier eine Zusammenfassung angewandt, um das Material auf die wesentlichen Inhalte zu reduzieren [Ma16]. Zusätzlich wurden die generalisierten Aussagen mit Hilfe der induktiven Kategorienbildung strukturiert. Im Gegensatz zur deduktiven Kategorienbildung, bei der die Kategorien auf theoretischen Vorüberlegungen basieren und durch Aussagen aus dem Material belegt werden sollen, ist beim induktiven Prozess das Kategoriensystem das Ergebnis der Analyse und wird für die weitere Interpretation der Fragestellung genutzt [Ma16].

3 Ausgewählte Ergebnisse der Studie

Im Ergebnis der durchgeführten Interviewstudie wurde ein Kategoriensystem entwickelt, das bei der Beantwortung der aufgestellten Forschungsfragen hilft und in den nächsten Abschnitten ausschnittweise erläutert wird. Zunächst wird dessen Aufbau überblicksartig beschrieben, bevor drei ausgewählte Kategorien näher beleuchtet werden.

3.1 Aufbau eines Kategoriensystems

Das erstellte Kategoriensystem baut auf den eingangs gestellten Forschungsfragen auf und dient somit gleichzeitig als Entscheidungsgrundlage für die Aufnahme einer Textstelle in eine (neue) Kategorie. Insgesamt wurden fünf Kategorien identifiziert: 1. *Gründe für die Prozessauswahl*, 2. *Änderung des Prozessablaufs*, 3. *Einflussfaktoren auf die Digitalisierung der Prozesse*, 4. *Erfolgreiche Einführung der digitalen Prozesse* und 5. *Auswirkungen der Digitalisierung*. Jede Kategorie wurde zusätzlich in die drei Unterkategorien *gesetzlich/politisch*, *organisatorisch* und *technisch* unterteilt. Ausnahmen bilden hier die Kategorie *Änderungen des Prozessablaufs*, bei der keine Aussagen im technischen Bereich identifiziert wurden, und die Kategorie *Erfolgreiche Einführung der digitalen Prozesse*, die mit dem organisatorischen Bereich nur eine Unterkategorie besitzt. Für eine bessere

Übersicht erfolgte zusätzlich eine weitere inhaltliche Unterteilung der Unterkategorien in Cluster. Es existieren jedoch auch Unterkategorien, bei der aufgrund der geringen Anzahl an zugeordneten Aussagen eine weitere Unterteilung nicht nötig war.

Die Vorstellungen der ausgewählten Kategorien enthalten jeweils eine tabellarische Zusammenfassung der Unterkategorien, welche die einzelnen Cluster und die Anzahl der verschiedenen Aussagen (Mehrfachnennungen wurden nicht gezählt) sowie exemplarisch eine der zugeordneten Aussagen (paraphrasiert und generalisiert) umfasst. In den Erläuterungen werden teilweise Originalzitate angegeben, wo diese besonders prägnant waren.

3.2 Gründe für die Prozessauswahl

In diese Kategorie wurden alle Aussagen aufgenommen, die zur Auswahl der zu digitalisierenden oder bereits digitalisierten Verwaltungsprozesse getroffen wurden. Das langfristige Ziel der Maßnahmen ist in der Regel die Digitalisierung aller Prozesse, jedoch kann dies nicht gleichzeitig geschehen und es wird zunächst eine Auswahl getroffen, um z.B. die notwendigen Voraussetzungen zu schaffen und zu lernen: „Das mit der Projektakte wollten wir einfach in unserem Bereich ausprobieren. Wir sollten es ja selber auch benutzen, was wir hier auf den Markt werfen." Diese Gründe und Einflussfaktoren sind in Tabelle 1 zusammengefasst.

Cluster (Anzahl Aussagen)	Aussagen (exemplarisch)
gesetzlich/politisch:	
Bundes-/Landesebene (3)	gesetzliche Vorgaben (z.B. Onlinezugangsgesetz)
Hochschulebene (2)	Wünsche/Prioritäten der Hausleitung
Prozessebene (4)	Prozesse mit benötigter Rechtssicherheit
organisatorisch:	
Zwänge/Druck (5)	Personalknappheit bzw. Personaleinsparung
Personenbezogene Gründe (5)	Vorschläge durch Dezernate und Fakultäten
Optimierung (8)	Standardisierung der Prozesse
Reichweite (6)	Prozesse mit großem Nutzen für Kunden
Projektmanagement (4)	Voraussetzungen und Situation müssen passen
Prozesseigenschaften (9)	keine bestimmten Prozesseigenschaften
technisch (4)	Fremdsysteme werden abgeschaltet

Tab. 1: Gründe für die Prozessauswahl überwiegen im organisatorischen Bereich.

Im gesetzlichen/politischen Bereich konnten Auswahlgründe auf drei verschiedenen Ebenen identifiziert werden. Auf *Bundes-/Landesebene* wurden gesetzliche Normen und Vorgaben angeführt, die einzuhalten sind und somit die Digitalisierung eines Prozesses erzwingen. Hierzu zählen auch europäische Richtlinien, die aber in der Regel durch nationales Recht umgesetzt werden [IT22]. Weitere Einflussfaktoren sind auf *Hochschulebene* v.a. die geäußerten Wünsche und Prioritäten durch die Leitung der Einrichtung, aber auch

der notwendige Abstimmungsbedarf. Zudem wurde auf *Prozessebene* die (Wieder-)Herstellung der Rechtssicherheit einzelner Verwaltungsverfahren als Auswahlgrund angeführt.

Dem organisatorischen Bereich sind die meisten der genannten Auswahlgründe zuzuordnen. Dazu zählen interne *Zwänge bzw. Druck*, dass der aktuelle analoge Prozess nicht mehr durchführbar ist oder Probleme erzeugt. Als *personenbezogene Gründe* wurden Vorschläge von verschiedenen Personenkreisen genannt, wie der Hochschulleitung oder den Fakultäten, aber auch von Personen mit Interesse und Zeit. Weitere Auswahlgründe lagen in der *Optimierung* der Prozesse, d.h. Standardisierung, Verbesserung der Effizienz, Erhöhung der Transparenz und Reduzierung von Kosten und Material, und ebenso in der *Reichweite* und damit der Sichtbarkeit für und auch die Auswirkung auf möglichst viele Personen. Hinzu kommen Einflüsse aus dem *Projektmanagement* in der Form, dass Personal für die Umsetzung vorhanden sein oder einfach die Situation passen muss. Konkrete *Prozesseigenschaften*, in Form von messbaren Merkmalen in Bezug auf Größe und Komplexität der Prozesse, wurden nicht genannt bzw. teilweise explizit ausgeschlossen. Jedoch wurde angeführt: „Ich würde jetzt tatsächlich mit diesen einfachen Sachen anfangen, wo relativ viele Nutzer oder Kunden der Verwaltung auch dranhängen und man eben auf diese Weise dann die Dinge auch beschleunigen kann." Hingegen wurden mitunter Kennzahlen wie Durchsatz oder auch Dauer des Vorgangs als Auswahlgrund angeführt.

Als Auswahlgründe aus technischer Sicht wurden lediglich von einigen Experten angeführt, dass eine Reaktion auf eine bevorstehende Abschaltung von Fremdsystemen benötigt oder die Digitalisierung eines Prozesses durch das Vorhandensein der dafür notwendigen Daten und Systeme begünstigt wurde.

3.3 Einflussfaktoren auf die Digitalisierung der Prozesse

In eine weitere Kategorie wurden alle Aussagen aufgenommen, die zu erkennbaren Einflüssen auf die Art und Weise der Digitalisierung von Verwaltungsprozessen getroffen wurden. Auch hier sind Faktoren aus allen drei Bereichen angeführt worden, welche in Tabelle 2 zusammengefasst sind.

Im gesetzlichen/politischen Bereich sind auf *Bundes-/Landesebene* auf der einen Seite wieder gesetzliche Rahmenbedingungen und Vorgaben zu nennen, die die Umsetzung beeinflussen, wobei nicht immer klar zu sein scheint, wie diese umzusetzen sind und teilweise auch Regelungen für die digitale Variante fehlen. Auf der anderen Seite besteht durch die landes- oder auch europaweiten Projekte ein Abstimmungsbedarf, aber auch ein Erfahrungsaustausch zwischen den verschiedenen Einrichtungen. Dieser bringt zwar Vorteile mit sich, scheint aber bei Forschungsgemeinschaften nicht immer gewünscht zu sein: „Wie gesagt, es geht nicht darum, dass die [Forschungsgemeinschaft] [eine Systemlösung] bekommt, das kriegen wir politisch nicht durchgesetzt." Auf *Hochschulebene* kann erneut

der Einfluss der Leitung der Einrichtung genannt werden, welche (teils mühsam) überzeugt werden muss. Hinzu kommen Mitbestimmungstatbestände, Schwerbehindertenvertretung und Gleichstellungsaspekt.

Cluster (Anzahl Aussagen)	Aussagen (exemplarisch)
gesetzlich/politisch:	
Bundes-/Landesebene (20)	Zusammenarbeit/Abstimmung von allen Einrichtungen im Land
Hochschulebene (11)	Mitbestimmungstatbestände
organisatorisch:	
Struktur der Einrichtung (8)	Vorgehen hängt von Aufbau der Einrichtung ab
Beteiligte (35)	Unterstützung durch zentrales Projektmanagement
Projektmanagement (38)	Erstellung einer Governance bzw. Digitalisierungsstrategie
Prozesse (22)	Prozesse müssen flexibel sein, Ausnahmen bzw. Sonderfälle müssen möglich sein
technisch:	
Systeme/Schnittstellen (29)	Dokumentenmanagementsystem (z.B. wegen rechtssicherer Datenaufbewahrung)
Anforderungen/Voraussetzungen/Maßnahmen (15)	Berechtigungsmanagement
Entscheidungen (10)	Vermeidung von Systemvielfalt wegen Komplexität und Wartbarkeit
Implementierungspartner (6)	Implementierungspartner sollte gesetzliche Rahmenbedingungen kennen

Tab. 2: Einflussfaktoren auf die Digitalisierung betreffen alle drei Teilbereiche.

Die Einflussfaktoren im organisatorischen Bereich bestehen aus vier Clustern mit teils auffallend vielen Nennungen. So hängt das Vorgehen von der *Struktur der Einrichtung* und damit auch den Entscheidungsstrukturen ab. Hinzu kommen die an der Umsetzung *beteiligten Personen*, d.h. Entscheidungsträger, Projektleitung, Personen mit fachlichem und technischem Wissen, aber auch Personen für rechtliche Fragen. Der Einbezug von Prozessbeteiligten ermöglicht zudem die Beachtung der Sicht der Nutzenden. Hilfreich sind ebenfalls Personen, die die Übersetzungsleistung zwischen Verwaltungs- und IT-Personal übernehmen können. Dabei wurde hervorgehoben, dass es wichtig aber leider nicht immer möglich ist, genügend und vor allem auch motiviertes Personal zusammenzustellen, das Zeit für die Umsetzung und auch für das Aneignen der neuen Technologien und Regeln hat. Betont wurde dabei, dass neues Personal und ein eventuell damit verbundener Generationenwechsel eine neue Denkweise mit sich bringen kann. Ein wichtiger Aspekt ist auch das *Projektmanagement* selbst, d.h. eine stabile Planung, Strategien und Maßnahmen mit klaren Zielsetzungen sowie das Erfassen aller relevanten externen und internen

Regelungen. Dabei muss darauf geachtet werden, dass eine Umstellung auf organisatorischer und technischer Ebene und eine Anpassung bisheriger Regelungen an den neuen digitalen Kontext notwendig ist. Dazu passend wurden insbesondere ein fehlendes professionelles Projektmanagement mit unklaren Strukturen, intransparentem Vorgehen und unterschiedlichen Zielstellungen sowie Termindruck und fehlende personelle Ressourcen als hinderlich empfunden. In Bezug auf die umzusetzenden *Prozesse* wurde ein flexibler Prozessablauf, der im Nachhinein noch anpassbar sein sollte, empfohlen. Vor allem im Verwaltungsbereich ist dabei eine Standardisierung möglich, nicht so in den Bereichen Studium/Lehre und Forschung: „Also je näher sie an den Kernprodukten dran sind, also Forschung und Lehre, da muss man auf die Individualitäten eingehen." Dabei ist eine Dokumentation der Prozesse in Form eines Praxishandbuchs eine gute Voraussetzung, da es sonst zu Problemen bei der Umsetzung kommen kann und der modellierte Sollprozess die Realität nicht widerspiegelt. Ein offenes Problem stellen jedoch Prozessausnahmen dar, die bei der Digitalisierung oftmals nicht berücksichtigt werden: „Und man vergisst dabei, dass die Ausnahmebehandlung gar kein Störfaktor ist, den man künstlich reinbringt, sondern einfach passiert."

Aus dem technischen Bereich wurden überwiegend die anzubindenden *Systeme/ Schnittstellen* als Einflussfaktoren benannt. Dabei handelt es sich um Fachverfahren, Campus-Management-, Identity-Management- und Forschungsinformationssysteme, welche sowohl Anwendungen externer Firmen als auch Eigenentwicklungen sind. Häufig wurde als zentrale Komponente ein Dokumentenmanagementsystem genannt, welches sich teilweise aber auch erst in Einführung befand. Die Einflussfaktoren sind dabei verschiedener Natur, z.B. fehlende Funktionalität in den Fachverfahren, Schnittstellen, die eine Unsicherheit in der Systemarchitektur sein können, und Umgang mit sensiblen Daten in „Fremdsystemen", aber auch ein notwendiger Parallelbetrieb bei der Modernisierung der Systemlandschaft bzw. zu erwartender Mehraufwand bei der Übernahme von Daten und Prozesse in neue Systeme. Hinzu kommen *Anforderungen/Voraussetzungen/Maßnahmen*, wie Berechtigungsmanagement, Sicherheits- und Datenschutzprinzipien und Erarbeitung und Umsetzung im Bereich System- und Datenarchitektur (auch für die Langzeitarchivierung). Hier wurden auch allgemeine Anforderungen angeführt, wie stabiler Betrieb und geringer technischer Unterstützungsaufwand. Problematisch ist jedoch immer noch die nicht vorhandene bzw. geklärte digitale Signatur. Der dritte Cluster umfasst alle getroffenen *Entscheidungen*. Hier wurden bspw. die Vermeidung von Systemvielfalt, die Wahl für bereits vorhandene Technologie bei der Umsetzung neuer Prozesse und auch die Vermeidung von Modifikationen oder Eigenprogrammierungen angegeben. Zum Teil wurden auch konkrete Umsetzungsstrategien genannt, d.h. standardisierte Prozesse mit hohen Fallzahlen können im Fachverfahren abgebildet werden, Akten und auch kollaborative Prozesse sollten hingegen im Dokumentenmanagementsystem durchgeführt werden. Einmal wurde auch die Verlagerung in die Cloud erwähnt. Teilweise wurden der Einfluss von *Implementierungspartnern* angesprochen, wobei vorwiegend Probleme bei der Absprache zwischen Einrichtungen und Dienstleistern sowie fehlendes Wissen der Dienstleister in Bezug auf gesetzliche Rahmenbedingungen benannt wurden. Zudem wurde auf Gefahren wie

Vendor-Lock-in und zu viel Einfluss des bereits bestehenden Implementierungspartners hingewiesen.

3.4 Erfolgreiche Einführung der digitalen Prozesse

Als dritte Kategorie soll exemplarisch vorgestellt werden, welche Aussagen über die Rahmenbedingungen einer erfolgreichen Einführung der digitalen Prozesse getroffen wurden. Dafür gab es lediglich Nennungen im organisatorischen Bereich, welche in Tabelle 3 dargestellt sind.

Cluster (Anzahl Aussagen)	Aussagen (exemplarisch)
organisatorisch:	
Kommunikation (14)	gute, transparente Kommunikation
Beteiligte (9)	motivierte Personen notwendig
Projektmanagement (14)	agile Ansätze
Prozesse (5)	Prozess optimieren/modernisieren

Tab. 3: Zur erfolgreichen Einführung digitaler Prozesse wurden organisatorische Faktoren benannt.

Einer der Hauptbereiche ist dabei „eine gute *Kommunikation* mit allen Beteiligten“, die Akzeptanz und Transparenz fördern und Verständnis schaffen sollte. Dies kann über das Sichtbarmachen von Vorteilen und Erfolgen geschehen, z.B. mittels konkreter Beispiele. Es müssen aber auch Probleme und Ängste der am Prozess teilnehmenden Personen ernst genommen und darauf eingegangen werden, sowie Erwartungshaltungen realisiert werden. Weitere Punkte sind, wie bei der Umsetzung, die *beteiligten Personen*, wo einerseits erneut Entscheidungsträger überzeugt, aber auch die am Prozess beteiligten Personen mitgenommen und vorbereitet werden müssen. Damit können Frustration und Abneigung vermieden und durch Schulungen kann auf die neuen Prozesse umgestellt werden. Gerade bei langjährigem Personal scheint hier die Umstellung schwierig zu sein und es kann auch zu Verweigerung kommen. Ein weiterer Aspekt mit vielen Nennungen ist das *Projektmanagement*. Neben der agilen oder iterativen Vorgehensweise wurde vor allem Erfahrung in verschiedenen Bereichen (wie dem Beschaffungs- oder Change-Management) genannt, aber auch Vorerfahrung aus anderen Projekten und externe Beratung. Zudem können Zwischenlösungen oder auch das Verschieben eines Projektes hilfreich sein. Als letzter Bereich wurden die *Prozesse* erwähnt. Als Erfolgsfaktoren wurden hier Prozessmodernisierung sowie die Gestaltung eines, soweit datenschutzrechtlich möglich, für alle Beteiligten transparenten Prozesses genannt. Dabei sollte vorab geklärt sein, welche Prozess(teil)e digitalisiert werden sollen bzw. können.

4 Diskussion und Ausblick

Zusammenfassend betrachtet, konnte aus der Interviewstudie ein multidimensionales Kriteriengeflecht sowohl bei der Auswahl der Prozesse als auch bei den Einflussfaktoren entwickelt werden, wobei über alle Kategorien hinweg vor allem Aussagen im organisatorischen und im politisch/gesetzlichen, jedoch weniger im technischen Bereich getroffen wurden. Deutlich wird das insbesondere in den Kategorien *Erfolgreiche Einführung der digitalen Prozesse* und *Änderung im Prozessablauf*, wo keine technische Unterkategorie existiert, aber auch in den anderen Kategorien überwiegen die organisatorischen Faktoren. In Bezug auf die eingangs gestellte Forschungsfrage und deren Unterfragen bedeutet dies, dass es weder eindeutige Hinweise auf Faktoren gibt, die ein zu digitalisierender Prozess haben muss oder nicht haben darf, noch treten die beabsichtigten Folgen der Digitalisierung stets ein.

Die Kategorie *Gründe für die Prozessauswahl* kann zur Beantwortung der ersten Unterfrage bzgl. der Prozesseigenschaften herangezogen werden und bestätigt die bereits gewonnene Erkenntnis, dass keine messbaren Prozesseigenschaften existieren [Le21a]. Vielmehr geht es um die Optimierung von teils problembehafteten Abläufen, aber auch um den Einflussbereich der Prozesse. Hinzu kommen äußere Einflüsse in Form von Gesetzen. Allerdings werden einige Aussagen (gerade in Bezug auf die Optimierung als Auswahlgrund) durch die Kategorie *Änderung des Prozessablaufs* teilweise widerlegt, da der Grundprozess oftmals (teils aus Gewohnheit) beibehalten wird oder die Optimierungsmöglichkeiten durch formale Vorgaben eingeschränkt sind. Zudem wurde hier deutlich, dass teilweise die gesetzlichen Grundlagen noch nicht geschaffen wurden.

Die Kategorie *Einflussfaktoren auf die Digitalisierun*g der Prozesse kann zur Betrachtung der zweiten Unterfrage genutzt werden. Im Gegensatz zu den anderen Kategorien wurden hier überdurchschnittlich viele technische Aussagen getroffen, wobei es zum Teil eine Beschreibung der vorhandenen Systemlandschaft war. Interessanterweise wurde diese teils als Argumentation für die Wahl der Technologie verwendet. Im organisatorischen Bereich ist besonders auffällig, dass oftmals das Fehlen von Einflussfaktoren, die positiv wahrgenommen werden, als störend empfunden wird und sich die Aussagen somit gegenseitig stützen. Auch wurden in diesem Bereich häufiger unkonventionelle Methoden benannt, wie das Liegenlassen von Projekten und das Warten auf einen Personalwechsel. Insgesamt entsteht hier zum Teil der Eindruck, dass einige Faktoren kaum steuerbar sind.

Die beschriebenen *Auswirkungen der Digitalisierung* und damit die Beantwortung der dritten Unterfrage umfassen vorrangig Vereinfachungen und Effizienzsteigerungen sowohl für die Beteiligten als auch in der Prozessoptimierung. Auffällig ist jedoch, dass hier teilweise ambivalente Aussagen bzgl. Zeit- und Personaleinsparungen existieren. Zudem wurden als negative Auswirkungen auch Probleme wie Kontroll- und Identitätsverlust benannt. Als mögliche Maßnahmen zum Umgang mit diesen Problemen konnten in der Kategorie *Erfolgreiche Einführung der digitalen Prozesse* primär eine gute Kommunikation und das Mitnehmen aller beteiligten Personen erfasst werden. Das stützt begleitende Erkenntnisse anderer Studien [KB21], dass eine als erfolgreich empfundene Digitalisierung

von Verwaltungsprozessen (nicht nur an Hochschulen) weniger von hoher Effizienz [LSL18], sondern vielmehr von breiter Akzeptanz und Transparenz [Le21b] geprägt ist. Verwundern mag allerdings, dass nur zweimal die Chance zur Modernisierung von Prozessen als Beitrag für die erfolgreiche Einführung der digital(isiert)en Prozesse angeführt wurde.

Als Limitation der Studie ist die Anzahl der durchgeführten Interviews zu nennen. Zwar war in den Aussagen eine Sättigung der ableitbaren Befunde erkennbar, doch ist nicht auszuschließen, dass die Auswahl der angefragten Personen subjektiv geprägt war und somit das Bild verzerrt haben könnte. Auch wurde die Codierung der Interviews nur durch eine Person durchgeführt, wodurch weitere subjektive Faktoren hinzugekommen sein mögen. Daher ist zu empfehlen, ausgehend von der beschriebenen Struktur und möglicherweise mit einem auf die bislang unscharf gebliebenen Aspekte vertiefenden Fokus, die Expertenbefragung zu wiederholen. Offen ist zudem die Ermittlung von Korrelationen zwischen den identifizierten begünstigenden und hemmenden Faktoren sowie möglichen Auswirkungen der Digitalisierung; hierfür sollte anschließend eine quantitative Studie durchgeführt werden.

Mit dem Vorliegen derartiger Erkenntnisse könnte die begonnene Kartographierung von Entscheidungsprozessen rund um die digitale Transformation der Hochschulverwaltung zu einem vertiefenden Lern- und Reflexionswerkzeug [Lu22] für IT-Entscheider weiterentwickelt werden. Damit könnten zudem die in der Hochschule stattfindenden Policy-Setzungen und -Implementierungen transparent gemacht [Al22] und partizipative Aspekte in Strategieprozessen weiter gestärkt [Lu20] werden.

Literatur

[Al22] Al Laban, F.; Engel, K.; Knoth, A.; Lucke, U.: Issues of Policy and Strategy for International Collaboration between European HEIs. to appear in Proc. EUNIS Annual Congress, 2022.

[Du05] Dunleavy, P., Margetts, H., Bastow, S., and Tinkler, J:. New Public Management Is Dead – Long Live Digital-Era Governance. Journal of Public Administration Research and Theory, 16(3), 2005, 467-494.

[Ha21] Hafer, J.; Kostädt, P.; Lucke, U.: Das Corona-Virus als Treiber der Digitalisierung? Eine kritische Analyse am Beispiel der Universität Potsdam. In: U. Dittler und C. Kreidl (Hrsg.): Wie Corona die Hochschullehre verändert, Springer Nature, 2021, S. 219-242.

[IT22] IT Planungsrat: Die Umsetzung des Onlinezugangsgesetzes (OZG) - Flächendeckende Digitalisierung der deutschen Verwaltung bis Ende 2022. https://www.it-planungsrat.de/foederale-zusammenarbeit/ozg-umsetzung

[KB21] Kulmann, S.; Bogumil, J.: Digitale Transformation in deutschen Kommunen. Das Beispiel der Bürgerämter und was man daraus lernen kann. In: Die Verwaltung, 54 (1), 2021, 1-27.

[LSL18] Lemcke S, Strickroth S, Lucke U: Effizienz in Zeiten der Digitalisierung: Schneller, besser, kostengünstiger? In: Zender R, Lucke U, Haase J, v.d. Heyde M, Leitner G, Meyer W (Hrsg): Lernen und Arbeiten im Wandel (LAiW) 2018, CEUR-WS, Bd. 2232.

[Le21a] Lemcke, S.: Komplexität und ihr Einfluss auf die Digitalisierbarkeit von Hochschulprozessen. In: Bolten-Bühler, R.; Dertinger, A.; Ellinger, D.; Thielsch, A.; Vanvinkenroye, J.; Zender, R. (Hrsg): Schöne neue (digitale) Welt?! Proc. Junges Forum Medien und Hochschulentwicklung (JFMH), Zenodo, 2021, 9-20.

[Le21b] Lemcke, S., Lehmann, A., Trommler, D., Lucke, U.: Akzeptanz und Effekte der Digitalisierung in der Hochschulverwaltung. HMD 58, 2021, 1020–1037.

[Le18] Lenk, K.: Verwaltungsinformatik in der Zeit nach dem E-Government. Verwaltung & Management 24(5), 2018, 217–225.

[Lo20] Loviscach, J.: Digitalisierung der Hochschullehre: Was wissen wir wirklich? In: R. Bauer et al. (Hrsg.), Vom E-Learning zur Digitalisierung – Mythen, Realitäten, Perspektiven, Münster: Waxmann, 2020, 84-100.

[Lu20] Lucke, U.; Hafer, J.; Hartmann, N.: Strategieentwicklung in der Hochschule als partizipativer Prozess. In: S. Goertz, B. Klages, D. Last, S. Strickroth (Hrsg.): Lehre und Lernen entwickeln – Eine Frage der Gestaltung von Übergängen. Potsdamer Beiträge zur Hochschulforschung 6. Universitätsverlag Potsdam: Potsdam, 2020, 99-118.

[Lu22] Lucke, U.; Köhler, T.; Assadi, G.; Moebert, T.; Müller, I.; Kranz, J.; Viehöver, W.; Böschen, S.; Manzeschke, A.: Interaktive Landkarten als ein Lernwerkzeug zur ethischen Exploration und Reflexion in Konstruktionsprozessen. erscheint in: Proc. DeLFI 2022, LNI P-322.

[Ma16] Mayring, P.: Einführung in die qualitative Sozialforschung. Eine Anleitung zu qualitativem Denken. Beltz, Weinheim, Basel, 2016.

[MD13] Margetts, H. and Dunleavy, P.: The second wave of digital-era governance: A quasi-paradigm for government on the Web. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 371(1987), 2013.

[PW14] Przyborski, A.; Wohlrab-Sahr, M.: Qualitative Sozialforschung. Ein Arbeitsbuch. Oldenbourg, München, 2014.

[RfII21] RfII – Rat für Informationsinfrastrukturen: Towards Common Scientific Data Infrastructures in Europe: Which Road to Take? Report on an International High Level Expert Workshop, Göttingen 2021. https://rfii.de/?p=7159

[vdH18] von der Heyde, M., Hartmann, A., Auth, G., Erfurth, C.: Zur disruptiven Digitalisierung von Hochschulforschung. Informatik Spektrum 41, 2018, 359–368.

Die Rolle der Informatik in der inter- und transdisziplinären Data Literacy Education

Eine empirische Studie der Fachperspektiven von Hochschullehrerinnen und Hochschullehrern zu Data Literacy

Julia Niemann-Lenz [1], Moritz Kreinsen [2] und Sören-Kristian Berger [3]

Abstract: Digitalisierung und Datafizierung in Gesellschaft, Politik, Wissenschaft und Wirtschaft haben unmittelbaren Einfluss auf die zukünftige Hochschullehre und darauf, welche Kompetenzen Studierende benötigen. Eine wesentliche Schlüsselkompetenz in diesem Diskurs bildet die *Data Literacy*. Dabei besteht Konsens, dass Lehr-Lern-Prozesse im Sinne einer *Data Literacy Education* (DLE) inter- bzw. transdisziplinär gestaltet werden müssen. Ungeklärt ist hierbei jedoch noch der Beitrag der Informatik zur DLE in der Hochschullehre. In einer Interviewstudie wird diese Problematik genauer untersucht, indem Hochschullehrende diverser Disziplinen u.a. zu ihrem Verständnis von DLE, dem Beitrag der Fachdisziplinen hierzu und den Anknüpfungspunkten zur Informatik befragt werden. Die bisherigen Ergebnisse zeigen Einigkeit der unterschiedlichen Disziplinen auf, dass die Stärkung der Data Literacy von Studierenden fächerübergreifend ein notwendiges und dringendes Anliegen ist. Weiterhin zeigt die Studie die Herausforderungen und die interdisziplinäre Diskrepanz im Aushandlungsprozess einer angemessenen Berücksichtigung von informatischen Inhalten in der hochschulweiten fachspezifischen DLE. Sie liefert aber auch Anknüpfungspunkte zur Überwindung dieser Diskrepanz.

Keywords: Data Literacy Education, Informatik, Hochschule, Lehre, Fachperspektiven, Interviews

1 Einleitung und Hintergrund

Die sich rasant weiterentwickelnden Informations- und Kommunikationstechnologien haben zu einer vernetzten und globalisierten Welt geführt, deren wichtigstes Gut Informationen und Daten sind. Im Zentrum dieses Wandels steht die Informatik und ist dabei Kern und Motor der Entwicklungen. Es ist ihre „Faszination, sich die Welt der Information und des symbolisierten Wissens zu erschließen und dienstbar zu machen. Informatik schafft neue Zugänge, neue Denkmodelle und zahllose automatisierte Helfer und Dienste" [GI06].

[1] Digital and Data Literacy in Teaching Lab, Zentrum für interdisziplinäre Studienangebote, Universität Hamburg, Jungiusstr. 11, 20355 Hamburg, julia.niemann-lenz@uni-hamburg.de, 0000-0003-1991-8535
[2] Digital and Data Literacy in Teaching Lab, Zentrum für interdisziplinäre Studienangebote, Universität Hamburg, Jungiusstr. 11, 20355 Hamburg, moritz.kreinsen@uni-hamburg.de, 0000-0002-2156-5946
[3] Digital and Data Literacy in Teaching Lab, Zentrum für interdisziplinäre Studienangebote, Universität Hamburg, Jungiusstr. 11, 20355 Hamburg, soeren-kristian.berger@uni-hamburg.de, 0000-0001-5147-2911

doi:10.18420/inf2022_121

Wenn über Daten im wissenschaftlichen Diskurs gesprochen wird, findet sich jedoch schnell eine breite Streuung an Disziplinen wieder, welche diese in unterschiedlichster Form, Art und Weise erheben, kontextuell aufbereiten und verarbeiten. Wenn über Methoden zur Datengewinnung und -verarbeitung gesprochen wird, dann beanspruchen ebenso viele verschiedene Disziplinen fachspezifische Kompetenz für sich. Jüngst besteht eine neue Wissenschaft, welche sich explizit auf die praktischen Methoden zur Arbeit mit Daten spezialisiert hat – die Data Science [GR18]. Diese Disziplin gilt als das „berufsspezifische Kompetenzprofil" [SBH19] der Informatik in Bezug auf das komplexe Thema *Daten* und liefert die Werkzeuge und Quellen zur praktischen Arbeit mit diesen, beispielsweise im Kontext von *Big Data*. In Abgrenzung hierzu kann *Data Literacy* (DL) als Fähigkeit verstanden werden, „die ein mündiger Bürger in der Digitalisierung benötigt, um sich in einem Überangebot von Daten und Informationen zurechtzufinden und fundierte Entscheidungen zu treffen" [SBH19]. Diese Fähigkeit ist disziplinenübergreifend relevant für Hochschulabsolvent:innen. Somit werden Kerninhalte aus der Informatik auch in anderen Fächern benötigt. „Die Informatik ist wie die Mathematik eine auf alle anderen Wissensgebiete ausstrahlende Grundlagen- und Formalwissenschaft. […] Ihre Produkte sind zwar überwiegend abstrakt, haben aber sehr konkrete Auswirkungen. Ihre Denkweisen dringen in alle anderen Wissenschaften ein" [GI06].

In der Frage nach der kritisch-reflexiven Auseinandersetzung und dem praktischen Umgang mit Daten in Lehr-Lern-Settings – der Data Literacy Education (DLE) – ist bislang noch ungeklärt, welche Inhalte und Methoden hierfür bestenfalls geeignet sind. Konsens besteht hingegen, dass dies interdisziplinär, d. h. in Zusammenarbeit der unterschiedlichen Disziplinen an einen spezifischen Lerngegenstand [Br20], erfolgen und grundlegende Kompetenzen in diesem Themenfeld fachübergreifend vermittelt werden müssen [SBH19]. Ridsdale et al. [Ri15] sehen bei diesem ambitionierten Vorhaben eine große Herausforderung in der Integration von DLE in bestehende Curricula. Zur möglichen Lösung dieses Problems sollen einige konkrete Inhalte und Kompetenzen direkt in existierende Fächer integriert werden. Hierbei muss jedoch berücksichtigt werden, dass sich nicht alle Lehrenden in der Lage sehen, die angesprochenen Kompetenzen zu vermitteln. Es stellt sich die Frage, welche Rolle folglich die Informatik bei dieser inter- und transdisziplinären DLE einnehmen kann.

Im Digitalisierungsdiskurs proklamiert die Informatik, sie „beschäftigt sich seit über 20 Jahren mit Lehrsystemen und Lehrszenarien für die Digitalisierung der Lehre im eigenen und in anderen Fächern, im engen interdisziplinären Austausch mit Bildungswissenschaftler[:innen] und Fachdidaktiker[:innen]. Entsprechende Forschungsergebnisse [...] bilden ein wesentliches Fundament für erfolgreiche digitalisierte Lehre" [Fa20]. Vor dem Hintergrund der zunehmenden Digitalisierung und Datafizierung der Gesellschaft entdecken auch viele andere Disziplinen informatische Methoden für sich. Der Bedarf an Wissenschaftler:innen, die eine Brücke zwischen Fachdisziplin und der Informatik schlagen können, ist genauso hoch wie der des außeruniversitären Arbeitsmarktes. Längst haben viele Hochschulen begonnen, diese Lücke durch spezielle Schnittstellenstudiengänge und DLE-Programme zu schließen

[EKR21]. Der Aushandlungsprozess zu einer gelungenen DLE hat aber gerade erst begonnen. Es gilt, die unterschiedlichen Perspektiven – einerseits der Informatik, andererseits vieler weiteren Fachdisziplinen – zu erforschen, um die inter- und transdisziplinären Prozesse in der Hochschullehre im Kontext von DL wirksamer zu gestalten.

2 Forschungsdesign

In einer empirischen Studie erforschen wir, welche Fachperspektiven Hochschullehrer:innen in Bezug auf die inter- und transdisziplinäre DLE besitzen, insbesondere zur Rolle der Informatik in diesem Diskurs. Die Fachperspektiven werden mithilfe von teilstandardisierten Leitfaden-Interviews erhoben und in einem interpretativen Prozess induktiv mit der qualitativen Inhaltsanalyse [Ma15] ausgewertet. Im Zentrum der Studie stehen folgende Fragestellungen:

- Welches Verständnis haben die Lehrenden unterschiedlicher Fachdisziplinen von Data Literacy?
- Welchen Beitrag sollte Data Literacy Education an einer Hochschule für die berufsbezogene Ausbildung und die Allgemeinbildung leisten?
- Welchen Beitrag können unterschiedliche Fachdisziplinen zur Data Literacy Education leisten?
- Welchen Stellenwert nimmt die Fachwissenschaft Informatik in Bezug auf die Zukunft des Lehrens und Lernens an der Hochschule ein?

Die Stichprobe in dieser Interviewstudie umfasst mind. 20 Hochschullehrer:innen aus verschiedenen Fach- respektive Arbeitsbereichen beispielsweise der Betriebswirtschaft, der Erziehungswissenschaft, den Geisteswissenschaften, der Informatik, der Medizin, der Rechtswissenschaft sowie den Sozial- und Wirtschaftswissenschaften.

3 Vorläufige Ergebnisse

Die Analysen der bisher geführten Interviews zeigen, dass die Interviewten die Stärkung der DL von Hochschulabsolvent:innen fächerübergreifend für ein notwendiges und dringendes Anliegen halten. DLE soll (1) zum souveränen Umgang mit den eigenen personenbezogenen Daten befähigen und (2) die berufliche Qualifikation verbessern und den Anforderungen des Arbeitsmarktes begegnen. Von Studierenden wird außerdem erwartet, dass sie (3) lernen die durch Digitalisierung und Datafizierung vorangetriebenen Veränderungen zu reflektieren und hierbei eine gestaltende Rolle einnehmen.

Der Begriff *Data Literacy* wird in den Disziplinen unterschiedlich ausgelegt. Insbesondere die interviewten Informatiker:innen betonen die technisch-praktische Seite der DL stark.

Sie beanspruchen die zugehörigen Kompetenzen und die Deutungshoheit in diesem Bereich für sich. Auch halten sie die Informatikstudierenden für *Data Literates*: „Unser Klientel im Informatikstudium ist ja schon so, dass die das tendenziell schon wissen" (Inf02). Die Interviewten anderer Fachdisziplinen halten die Rolle der Informatik hingegen für „meistens überbewertet" (BWL01). Sie verweisen auf die Kontextgebundenheit von Daten (Med01), die eigene Fachtradition und die Bedeutung von nicht-informatischem Wissen und Techniken (z.B. Statistik, Erhebungsmethoden, Interpretation und Ableitung von Handlungsempfehlungen). Da solche *fachspezifischen Datenkompetenzen* für die bisher interviewten Nicht-Informatiker:innen im Vordergrund stehen, verwundert es nicht, dass sie auch kaum die Notwendigkeit sehen, Methodenwissen aus der Informatik oder Programmierkenntnisse stärker in ihre jeweiligen Studienprogramme zu integrieren. Allerdings stehen unsere Interviews mit *datennahen* Disziplinen zum jetzigen Zeitpunkt noch aus. Es bleibt abzuwarten, ob Lehrende bzw. Fächer, die bereits stärkere Schnittmengen mit der Informatik haben, dieser Meinung folgen (z.B. Computational Social Science oder Digital Humanities).

Gleichwohl: Die bisher interviewten Informatiker:innen beurteilen dies durch die Bank diametral gegensätzlich und sehen einen erhöhten Bedarf nach der Informatik-Perspektive in der Lehre anderer Fächer. Zwar muss nicht jede:r Studierende programmieren lernen, jedoch sollten die notwendigen Grundlagen vermittelt werden, um Studierende anderer Disziplinen *sprachfähig* zu machen: „Muss jeder Mensch programmieren lernen? Ich würde sagen nein. Aber muss jeder Mensch mal diese Kontrollstrukturen gesehen haben, um eine grobe Vorstellung davon zu entwickeln, was Programmieren ist? Ich würde sagen ja" [Inf02]. Diese technisch-praktische Komponente der DLE ist dabei mit der reflexiv-kritischen verwoben. Inf01 betont, dass es schwierig ist, hier ein gesundes Maß zu finden: „[Es] ist wichtig, dass man nicht alles blind glaubt, aber es ist auch wichtig, dass man nicht alles hinterfragt". Unklar bleibt jedoch, wer derartige Lehre übernehmen kann und sollte. Inf02 formuliert die Sorge, dass aus der Informatik heraus in der Zukunft sehr viel „Servicelehre" zu leisten ist. Er sieht die Verantwortung für diese Lehrinhalte bei den einzelnen Fächern. Inf03 plädiert für die Notwendigkeit von Schnittstellenprofessuren, die einerseits disziplinär verankert sind, andererseits die fachliche Brücke zur Informatik schlagen können. Auch in transdisziplinären Lehrtandems sieht sie eine große Chance.

4 Zwischenfazit und Ausblick

Die Interviews zeigen bereits jetzt die Bandbreite des DL-Verständnisses unterschiedlicher Fachdisziplinen und ihren variierenden Zugang zu DLE auf. Weiterhin zeigt die Studie die Herausforderungen und die interdisziplinäre Diskrepanz im Aushandlungsprozess einer angemessenen Berücksichtigung von informatischen Inhalten in der hochschulweiten fachspezifischen DLE. Sie liefert aber auch Anknüpfungspunkte zur Überwindung dieser Diskrepanz. Wir planen die Studie im Sommer 2022 abzuschließen und sind der Meinung, dass sie einen integrativen Beitrag zum Dialog über eine zukunftsorientierte DL-Ausbildung an der Hochschullehre leisten wird.

5 Literatur

[Br20] Braßler, M.: Praxishandbuch Interdisziplinäres Lehren und Lernen. 50 Methoden für die Hochschullehre, 1. Auflage, Beltz Juventa, Weinheim Basel, 2020.

[EKR21] Ebeling, J., Koch, H., Roth-Grigori, A.: Data Literacy Education an deutschen Hochschulen. Kompetenzen im kritischen Umgang mit Daten, Stifterverband, Essen, 2021.

[Fa20] Fakultätentag Informatik: Positionspapier "Digitalisierung der Hochschullehre". Abgerufen von https://www.ft-informatik.de/pdf/FTI-2020-Position-Digitalisierung-Hochschullehre-200327.pdf. 2020.

[GI06] GI, Gesellschaft für Informatik e.V., “Was ist Informatik? Unser Positionspapier”. Abgerufen von www.gi.de/fileadmin/GI/Hauptseite/Themen/was-ist-informatik-lang.pdf, 2006.

[GR18] Grillenberger, A. & Romeike, R.: Was ist Data Science? Ermittlung der informatischen Inhalte durch Analyse von Studienangeboten. In N. Bergner, R. Röpke, U. Schroeder & D. Krömker (Hrsg.): Hochschuldidaktik der Informatik HDI 2018. 8. Fachtagung des GI-Fachbereichs Informatik und Ausbildung/Didaktik der Informatik. 12.-13. September 2018 an der Goethe-Universität Frankfurt am Main. Universitätsverlag Potsdam, Potsdam, 119-134, 2018.

[Ma15] Mayring, P.: Qualitative Inhaltsanalyse: Grundlagen und Techniken, 12. überarbeitete Auflage, Beltz, Weinheim Basel, 2015.

[Ri15] Ridsdale et al.: Strategies and Best Practices for Data Literacy Education: Knowledge Synthesis Report, 2015.

[SBH19] Schüller, K., Busch, P., Hindinger, C.: Future Skills: Ein Framework für Data Literacy – Kompetenzrahmen und Forschungsbericht. Arbeitspapier Nr. 47, Hochschulforum Digitalisierung, Berlin, 2019.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1417

bwLehrpool als offene und flexible Betriebsplattform für PC-Pools mit optionalem Fernzugriff

Status und Entwicklung eines kollaborativ genutzten Dienstes für Hochschulen

Dirk von Suchodoletz[1], Simon Rettberg[1], Manuel Bentele[1], Steffen Ritter[2] und Christian Rößler[1]

Abstract: Moderne Hochschul-IT stellt flexible IT-Infrastrukturen auf dem Campus bereit, die hybride Lehr- und Lernszenarien auf Basis vielfältiger Softwareumgebungen erlaubt. bwLehrpool ist eine über mehrere Jahre kollaborativ entwickelte Softwareplattform zum aufwandsarmen Betrieb von PC-Pools mittels bereitgestellter Desktop-Virtualisierung. Letztere erlaubt kollaboratives Erstellen und hochschulübergreifendes Teilen kuratierter Lehrinhalte. Diese können seit einiger Zeit über Campusgrenzen hinweg im Fernzugriff angeboten werden, was einen Weiterbetrieb in Zeiten der Pandemie und eine bessere Auslastung der bestehenden Ressourcen erlaubt. Ebenso wird bwLehrpool als rechtssichere Basisplattform für E-Prüfungen im Zusammenhang mit Learning Management Systemen (LMS) genutzt. Das modulare Konzept bietet neben einfacher Einbindung in bestehende Services eine übergreifende Steuerung und Überwachung großer Rechnerinstallationen. bwLehrpool wird damit zu einem Baustein zeitgemäßer IT-Strategien von Hochschulen, die auf Bildungsgerechtigkeit und Digitale Souveränität Wert legen.

Keywords: PC-Pool-Steuerung, Desktop Virtualisierung, digitale Lernumgebung, Administration, Ressourcen Monitoring, moderne IT-Infrastrukturen, Netzwerk-Boot Linux

1 Motivation

Die Erwartungen an moderne und flexible IT-Strukturen für die Lehre stellen Hochschulen vor neue Herausforderungen. Der signifikant gestiegene Umfang und die Ausdifferenzierung erwarteter Softwareumgebungen erfordern neue Ansätze. Die im Laufe der Zeit entstandenen dezentralen IT-Strukturen gerade im Bereich der Desktop-Versorgung in PC-Pools generieren hohen Administrationsaufwand und zeigen zunehmend Defizite in der IT-Sicherheit oder der rechtskonformen Durchführung elektronischer Prüfungen. Vielfach ist die derzeitige Situation durch Softwareumgebungen geprägt, die im Verteilverfahren persistent und statisch konfiguriert auf definierte Desktop-Rechner installiert werden. Nach einer Auswahl der Pakete durch Lehrpersonen entscheiden am Ende Administratoren, wie Umgebungen aussehen und wann neue Versionen und Pakete verfügbar sind. Dieses Verfahren ist oft unflexibel, da

[1] Albert-Ludwigs-Universität Freiburg, Fahnenbergplatz, 79085 Freiburg im Breisgau, dirk.von.suchodoletz@rz.uni-freiburg.de, simon.rettberg@rz.uni-freiburg.de, manuel.bentele@rz.uni-freiburg.de, christian.roessler@rz.uni-freiburg.de
[2] Hochschule Offenburg, Badstraße 24, 77652 Offenburg, steffen.ritter@hs-offenburg

doi:10.18420/inf2022_122

sehr lange Änderungszyklen (Aktualisierung z.B. nur zu Semesterbeginn) dominieren. Abstimmungsschwierigkeiten und Missverständnisse zwischen Administratoren und Lehrenden sind an der Tagesordnung. Gründe dafür sind beispielsweise falsche Versionen, unvollständige Pakete, aufgeblasene Installationen, zu wenige/viele/falsche Beispiele für bestimmte Softwarepakete sowie ungünstige Konfigurationen und Voreinstellungen. Vielfach bestehen Einschränkungen in der Auswahl durch Festlegung auf nur ein Softwarepaket, weil es zu aufwändig erscheint, weitere beziehungsweise alternative Produkte zu pflegen. Verschärft wird die Situation durch einen Mangel qualifizierten Personals in der dezentralen IT-Administration. Ähnliches gilt für das Basis-Betriebssystem solcher Umgebungen. Dies schränkt Lehrkräfte in ihren Wahlmöglichkeiten ein und verhindert neue Ansätze in der Lehre. Hinzu kommen organisatorische Inflexibilitäten, die verhindern, dass bestehende IT-Lehrräume optimal ausgelastet werden können.

Neben der flexiblen Bereitstellung moderner Lehr- und Prüfungsformate soll eine bessere Kollaboration zwischen Lehrenden auch unterschiedlicher Hochschulen durch Austausch von Inhalten und übergreifende Kooperation erreicht werden. Ebenso wünschen viele Hochschulen, mehr hybride Lehr- und Lernszenarien anzubieten, und bestehende Campus-Ressourcen auch aus der Ferne nutzen zu können [KD21]. Das erlaubt, in Randzeiten häufig brachliegende Kapazitäten und Softwareangebote der PC-Pools Studierenden unabhängig ihrer persönlichen Ausstattung zugänglich zu machen.

2 bwLehrpool als modulare Softwareplattform für die Lehre

bwLehrpool schafft eine Plattform zum campusweiten Betrieb von PC-Pools für den Betrieb klassischer Linux- oder Windows-Umgebungen. Es erlaubt neben PC-Pools an beliebigen Standorten die Bereitstellung flexibler Kiosk-Systeme in großer Zahl (Bibliotheksterminals, Self-Service-Stationen) und die Auslieferung standardisierter Arbeitsplätze in Forschungsseminaren. Das für den Lehr- und E-Klausuren-Betrieb notwendige Basissystem wird durch eine schlanke, zentrale Infrastruktur bereitgestellt [Ri16]. Diese setzt auf klassische Technik zum Starten eines PCs über das lokale Netzwerk. Das Basissystem wird beim Hochfahren passend konfiguriert und bietet die Grundlage für eine umfangreiche Virtualisierung. Letztere kann durch Containerisierung oder diverser Hypervisoren ein breites Spektrum verschiedener kollaborativer Lehr-, Lern- und Forschungsumgebungen anbieten. bwLehrpool gestattet auf diese Weise einen leichtgewichtigen und beliebig skalierenden Ansatz parallel zu bestehenden Betriebsmodellen ohne lokale Installation. Die Nutzung von Linux als Basisbetriebssystem stellt die Unterstützung einer breiten Hardwarelandschaft sicher.

Die administrative Trennung des Hardwarebetriebs der Rechner an verschiedenen Standorten durch die jeweilige zuständige Einrichtung, Bereitstellung der Basisumgebung durch das Rechenzentrum sowie Pflege der Softwareumgebungen in Form virtueller Maschinen mit unterschiedlicher Ausstattung durch Lehrende ist ein zentrales Merkmal

der Plattform. Dieser Ansatz und weitere Module zum standardisierten Austausch mit weiteren Systemen sorgt für eine Vereinfachung des IT-Managements von PC-Pools und weiteren Maschinen im Netzbootbetrieb [Ba19]. An jedem Standort ist ein sogenannter Satellitenserver das administrative Kernstück unter der Kontrolle der jeweiligen Einrichtung. Für baden-württembergische Hochschulen wird zusätzlich ein zentraler Masterserver betrieben, um insbesondere den Austausch kuratierter Umgebungen über Hochschulgrenzen hinweg zu befördern (Abb. 1).

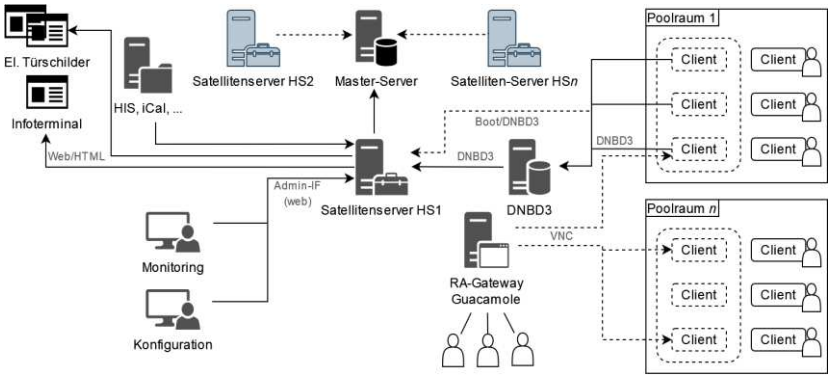


Abb. 1: Zentrale Module der bwLehrpool-Plattform mit dem im gewissen Umfang mandantenfähigen Satellitenserver im Zentrum der Steuerung

bwLehrpool ist auf vielfältige Weise darauf ausgelegt, sich reibungslos in bestehende PC-Systemlandschaften von Hochschulen zu integrieren, indem es sowohl verschiedene Formen der User-Authentifizierung, von Homeverzeichnissen und der Einbindung weiterer Ressourcen, wie Drucker bietet. Für die Nutzung in der Lehre werden eigens erstellte Werkzeuge für die Erstellung und Pflege von Softwareumgebungen in Virtuellen Maschinen (VMs), zur Auswahl dieser Umgebungen am Desktop in den PC-Pools, zum Anfahren der gewählten VM, zur Einbindung von Beamern oder der Konfiguration von Multi-Monitor-Setups sowie der Anbindung von Kalendern genutzt. Deshalb benötigt ein Rollout des Dienstes an einer zusätzlichen Hochschule inklusive Basisschulung selten mehr als einen Tag.

3 Standard-Linux mit Erweiterungen

bwLehrpool setzt das Konzept eines zustandslosen Betriebs auf zwei Ebenen um. Zum einen bietet es die Grundlage für einen dezentralen, netzwerk-bootenden Linux-Desktop, und zum anderen eine breite Auswahl darauf aufsetzender Virtualisierungsoptionen. Bei Stateless Remote-Boot handelt es sich um ein im Unix-Umfeld seit mehr als drei Jahrzehnten angewandtes Bootverfahren zum schnellen und aufwandsarmen Betrieb größerer Rechnerzahlen. Im Bootverlauf ausgehend von UEFI/iPXE setzt bwLehrpool auf die üblichen Linux-Mechanismen für die automatische Erkennung von Hard- und

Software bei jedem Start einer Maschine. Eine persistente Konfiguration kann auf den Ebenen eines Campus, von Räumen oder einzelnen Rechnern im Satellitenserver zentral hinterlegt werden. Das Basis-Linux wird durch ein eigens entwickeltes Build-System automatisiert erstellt. Als Basis für existierende (Binär-)Pakete dienen die Paketquellen der jeweils aktuellen Ubuntu LTS-Version. Das Generieren des Systems (Betriebssystemkern mit Ausgangsdateisystem) läuft dabei komplett automatisiert in einem systemd-nspawn-Container ab.

Nicht nur zum Einsatz in bwLehrpool wurde ein spezielles Network Block-Device (DNBD3) für den Linux-Kernel implementiert [RSB19]. Es dient dazu, das Hauptdateisystem des Linux-Grundsystems sowie VM-Images[3] auszuliefern. Dateien von auf dieser Weise bereitgestellten Dateisystemen werden blockweise gelesen. Dies bringt den Vorteil, dass nur die zur Laufzeit der einzelnen Maschinen oder VMs tatsächlich benötigten Blöcke transferiert werden, jedoch nicht zwingend der Inhalt einer vollständigen Installation. In Kombination mit einer Nur-Lese-Bereitstellung erlaubt dieser Mechanismus ein einfaches Caching häufig angefragter Blöcke bei sehr effizienter Übertragung, die durch Kompression weiter verbessert werden kann. Die Leseperformance kann so auch bei großen Maschinenzahlen oder in unterdimensionierten Segmenten eines Netzwerks hochgehalten werden.

Zur Verbesserung des Antwortverhaltens und Erreichung einer hohen Verfügbarkeit können PC-Clients mehrere DNBD3-Server gleichzeitig einbinden. Der am schnellsten antwortende Server kommt als primäre Quelle zum Einsatz. Sollte dieser aus Sicht des Clients nicht oder nur verzögert reagieren, wählt der Client einen alternativen Server aus der konfigurierten Liste, was für hohe Redundanz und Manipulationssicherheit sorgt. Die in DNBD3 eingebaute Versionierung ermöglicht nahtloses Aktualisieren ausgelieferter Dateisysteme und VM-Images. Eine lokale Installation der Betriebssysteme auf Arbeitsplatzrechnern wird auf diese Weise vermieden.

4 Desktop-Virtualisierung und Remote-Zugriff

Neben dem Linux-Grundsystem, welches je nach Konfiguration bereits einen umfangreichen grafischen Desktop anbietet, wird das System durch den Einsatz verschiedener Hypervisoren (VMware, VirtualBox, QEMU/KVM) und Containerisierung (Docker) sehr flexibel und unabhängig von der Basisadministration, weitgehend durch die Lehrenden, anpassbar. Das gestattet eine strikte Trennung von Zuständigkeiten mit starker Delegationskomponente, wodurch Lehrende nicht mehr auf die Erreichbarkeit der Administratoren angewiesen sind und somit zeitlich und örtlich unabhängig ihre gewünschten Softwareumgebungen anpassen können.

Funktionen zur Bereitstellung lokal verfügbarer Ressourcen mittels Gerätedurchreichung erlauben in der jüngsten Version die Verwendung lokaler Geräte (z.B. Grafikkarten) in

[3] Ein VM-Image bezeichnet ein zur VM zugehöriges Festplattenabbild

einer virtualisierten Softwareumgebung (Container, VM). Dadurch lassen sich eine Vielzahl an Anwendungsszenarien, insbesondere in den Bereichen CAD und KI, deutlich beschleunigen oder überhaupt erst umsetzen, indem nativ vorhandene Hardware direkt in der VM bzw. dem Container verwendet werden kann.

Zudem können Pool-Clients durch einen auf Ebene des Basissystems ansetzenden Remote-Zugriff ortsunabhängig genutzt werden. Mit Hilfe eines Webbrowsers auf beliebiger Plattform können Studierende und Lehrende von Zuhause aus auf hochschuleigene Poolrechner zugreifen und diese nutzen, als wären sie vor Ort. Besondere Leistungsanforderungen werden an die Maschinen der Nutzenden nicht gestellt. Bestehende Lizenzserver können problemlos verwendet werden, da alle Programme direkt auf den Rechnern innerhalb der Hochschulen ausgeführt werden.

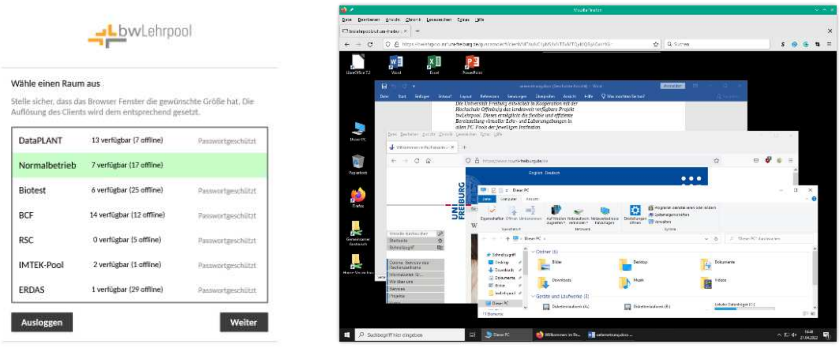


Abb. 2: bwLehrpool-Remote für den Fernzugriff im Web-Browser - Auswahl des virtuellen Raums (links) und gestartete Session nach Auswahl aus Liste (rechts)

Die Zuteilung einzelner Maschinen erfolgt über einen zentralen Eingangsproxy durch Auswahl eines virtuellen Pools oder einer Rechnergruppe mit anschließender Initialisierung des Remote-Desktops via Apache Guacamole[4]. Eine einfache Abgrenzung dieser Gruppen ist vergleichbar zum physischen Zugriff durch ein geteiltes Passwort möglich. Somit lässt sich verhindern, dass leistungsfähige, spezialisierte Rechner durch unbedarfte Nutzende mit trivialen Workloads blockiert werden. Die Zahl der belegten Rechner wird im Hintergrund überwacht, um mittels Wake-on-LAN dafür zu sorgen, dass stets eine einstellbare Anzahl freier Plätze bereitsteht, um Nutzenden einen sofortigen Sitzungsstart zu ermöglichen und gleichzeitig den Energieverbrauch durch unnötig eingeschaltete Computer zu minimieren.

[4] https://guacamole.apache.org

5 Vereinfachte IT-Administration

Die bwLehrpool-Plattform beinhaltet ein umfangreiches Monitoring, das über das Admin-Interface des Satellitenservers zugänglich ist. Hierzu stellt die Weboberfläche eine Vielzahl von Statistiken und Informationen über konfigurierte Rechnerpools zur Verfügung. So gibt es ein Modul zur Visualisierung der Auslastung der mit bwLehrpool gesteuerten Pools. Dies betrifft die Nutzung und Auslastung vorhandener Rechner, die sich aggregiert auf Räume oder heruntergebrochen auf einzelne Clients betrachten lässt. Über die verwendete Hardware werden ebenfalls Daten erhoben, sodass sich für jeden Client Details u.a. über verbaute Festplatten oder SSDs, RAM-Riegel, Mainboard-Typ und BIOS bzw. UEFI-Revision anzeigen lassen. Die Verteilung verschiedener Hardware-Merkmale oder Rechnertypen lässt sich über Tortendiagramme visualisieren.

Außerdem werden aktuelle Betriebsparameter wie CPU-Temperatur oder RAM-Auslastung bereitgestellt. Über eine Suchfunktion lässt sich ausgeben, wie viele Rechner eines bestimmten Modells im Einsatz sind, und in welchen Räumen sie sich befinden. Eine bevorstehende Ablösung eines alten Rechnermodells oder das Erweitern der Maschinen mit weniger als n GB RAM lässt sich somit komfortabel planen.



Abb. 3: Übersicht von Pool-PCs im Zeitverlauf und Visualisierung von Verteilungen

Zusätzlich zu grafisch aufbereiteten Statistiken bietet bwLehrpool außerdem ein zentralisiertes Logging, das über Client-Ereignisse informiert. Darunter fallen etwa Sitzungsstart und -ende, auftretende Fehler in Soft- und Hardware, falls z.B. Authentifizierungsserver nicht erreicht oder VM-Images nicht eingebunden werden können; ebenso Hardwarefehler, die an Pool-PCs auftreten.

Ein weiteres Modul ist die Anzeige von Raumbelegungen in Kombination mit konfigurierbarer Kalenderanzeige. Dieser Mechanismus bietet nach serverseitiger Einrichtung eines Raumplanes, eine vielfältig konfigurierbare Visualisierung der tatsächlichen Nutzung eines Raumes. Dabei werden Zustände, wie „länger nicht gesehen“, ausgeschaltet, ungenutzt, remote oder lokal belegt, farblich unterschieden. Die Visualisierungen werden als einfache Webseite exportiert und können mit üblichen Browsern angezeigt werden. Günstige Tabletcomputer oder E-Ink-Anzeigegeräte lassen

sich so ohne Bindung an proprietäre Herstellersoftware als elektronische Türschilder verwenden. Das Modul erlaubt, nach Einbindung einer Kalenderdatenquelle, einen raumbezüglichen Stundenplan anzuzeigen. Als Datenquellen können derzeit HISinOne, iCal-Dateien, Microsoft Exchange und Davinci eingebunden werden.

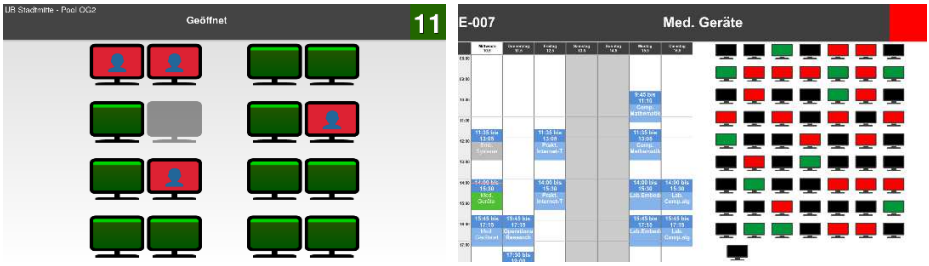


Abb. 4: Einfache Raumanzeige mit aktueller Belegung (links) bzw. Anzeige mit Kalender (rechts)

6 Unterstützung der Lehre

Ausgehend von fachspezifisch orientierten Angeboten aus reinen Softwarepaketen lassen sich themen-, fach- und nutzerorientierte Open Educational Resources (OER) anlegen, die aufgabenspezifisch mit Beispielen, Dokumentation, Erklärvideos, Linksammlungen etc. angereichert werden [CB17]. Diese Umgebungen transportieren den Ansatz der analogen Welt - der Aufbereitung eines Themas für die Lehre - in die digitale Welt. So vorgefertigte Umgebungen reichen weit über den rein technischen Aspekt hinaus. Um einen Austausch und eine leichte Weiterentwicklung der Inhalte bei gleichzeitiger Versionierung und Spezialisierung zu erlauben, wird mit bwLehrpool eine Plattform bereitgestellt, die Versionierung und Qualitätssicherung unterstützt und fördert. Dies ermöglicht Lehrenden, im Team hochschul- und hochschulartenübergreifend themenorientierte OERs in Form virtueller Lehr- und Arbeitsumgebungen zu erschaffen, die themenbezogen in Kursen eingesetzt, zum zeitlich freien Lehren und Lernen verwendet und mit Studierenden geteilt werden können. Die gegebenen Freiheitsgrade beim Einsatz von Software und Konfigurationen erlaubt eine optimale Nutzung und Weitergabe vorhandenen Expertenwissens.

Gleichzeitig können diese Umgebungen helfen, Studierende an moderne Forschungs- und Arbeitsmethoden heranzuführen. Wie in anderen Arbeitsgebieten kann so durch Kollaboration schnell eine hohe Qualität der Inhalte erreicht werden, die Einzelne allein nicht leicht erreichen könnten. Solche Sammlungen an Inhalten werden im Laufe der Zeit wachsen, so dass die Auswahl größer wird und dadurch neue Ideen für Lehre und Lernen befördert werden.

Dieser kollaborative Ansatz setzt voraus, Lehrenden die Erstellung, Verwaltung und Verwendung benötigter Lehrumgebungen weitgehend zu vereinfachen. Ein wesentlicher Aspekt besteht daher neben der Bereitstellung einer umfangreichen

Anwendungsdokumentation[5] in der bwLehrpool-Suite. Dieses Dienstprogramm wurde in der systemübergreifenden Sprache Java geschrieben, um auf vielen Plattformen mit unterschiedlichen Betriebssystemen lauffähig zu sein.

Um Lehrenden die Erstellung eigener Umgebungen zu erleichtern, werden von den einzelnen Hochschulen und Universitäten vorgefertigte Standardvorlagen in Form von angepassten VM-Images angeboten, in denen übliche Standardsoftware enthalten und vorkonfiguriert ist. Diese Standardvorlagen können vom Lehrpersonal bequem und ohne weiteres für den gewünschten Lernzweck erweitert werden (z.B. durch Installation von gewünschten Softwarepaketen).

6.1 E-Prüfungen

Zur Konfiguration und Ausführung elektronischer Prüfungen wurde ein eigenes Modul entwickelt. Im Prüfungsmodus gestartete Rechner erhalten ein zusätzlich abgesichertes Grundsystem, das VMs in ihrer Nutzung einschränkt, dafür sorgt, dass nur bestimmte Prüfungsumgebungen gestartet werden können oder eine ausgewählte Veranstaltung direkt startet. Zusätzlich lassen sich weitere Beschränkungen festlegen; so können etwa USB-Ports deaktiviert und die Netzwerkkommunikation eingeschränkt oder ganz deaktiviert werden. Üblicherweise werden ganze PC-Pools in den Prüfungsmodus versetzt; die Umschaltung in und aus dem Prüfungsmodus kann zeitlich anhand des serverseitigen Prüfungskalenders im Admin-Interface des Satellitenservers vorgegeben werden und erfolgt in Minutenschnelle.

Im Verbundprojekt „Partnerschaft für innovative E-Prüfungen Projektverbund der baden-württembergischen Universitäten“ (PePP)[6] ist die Weiterentwicklung von bwLehrpool Teil eines sogenannter Reallabore, in denen innovative E-Prüfungslösungen technisch weiterentwickelt und erprobt werden. Im Rahmen des Projektes soll bwLehrpool als Prüfungs-Grundsystem ausgebaut und die an der Universität Freiburg bereits realisierten E-Prüfungsszenarien an andere Universitäten transferiert und flexibilisiert werden. Dazu werden zwei neue Entwicklungsvorhaben verfolgt, die mit den weiteren Maßnahmen des Verbundprojekts verschränkt sind und auf existierenden Vorarbeiten aufbauen. Ein Vorhaben verfolgt die Entwicklung einer Virtual Desktop Infrastructure (VDI) für bwLehrpool, welche virtualisierte Prüfungsumgebungen auf bestehenden Cloud Infrastrukturen (z.B. der bwCloud[7]), bereitstellen soll. So können praxisnahe Klausurformen mit hohen Teilnehmendenzahlen erprobt werden. Weiter werden bestehende Möglichkeiten erweitert, von entfernten Rechnern aus per Browser auf die abgesicherte Prüfungsumgebung zuzugreifen. Dabei wird zusätzlich evaluiert, mobile Geräte wie beispielsweise Chromebooks oder Notebooks im Besitz der Studierenden

[5] https://www.bwlehrpool.de/wiki
[6] https://www.hnd-bw.de/projekte/pepp
[7] https://www.bw-cloud.org

(sogenanntes Bring-Your-Own-Device) einzubinden. Hierbei ergeben sich große Synergien hinsichtlich Flexibilisierung und Mobilisierung elektronischer Prüfungen.

Im Hinblick auf die gestiegene Nachfrage nach Proctoring-Systemen[8] soll die vorhandene Lösung zur Steuerung der Pool-Bildschirme (Pool-Video-Switch) auch für bwLehrpool-Remote nutzbar gemacht und um weitere Elemente (z.B. webRTC-Komponenten) ergänzt werden, um so eine virtuelle Prüfungsaufsicht zu erproben. Vorrangig wird ein „Whitebook - Dos & Don'ts" zum Proctoring mit bwLehrpool entwickelt, in dem in Zusammenarbeit mit dem PePP-Querschnittscluster "Recht und Akzeptanz" auf technische und rechtliche Herausforderungen eingegangen wird.

7 Laufende Entwicklungen

Aktuelle Entwicklungen umfassen u.a. die Optimierung der Durchreichung lokaler PCI-Geräte (beispielsweise Grafik- oder Ein-/Ausgabekarten) an virtualisierte Lehrumgebungen, um intensive Berechnungen zu beschleunigen. Dabei soll dies auch für die Nutzung in Containern ermöglicht werden. Diese zunehmend komplexen Module erfordern, ebenso wie die Einbindung proprietärer Komponenten (z.B. spezielle Grafiktreiber oder kommerzielle Hypervisoren), umfangreiches und aufwändiges Testen, weshalb schrittweise Continuous Integration (CI) und Continuous Deployment (CD) ausgebaut werden soll.

In einem weiteren Schritt soll eine Web-Anwendung zur Bearbeitung von VMs etabliert werden, die bwLehrpool-Remote nutzt und damit einen Transfer zunehmend größerer VM-Images über vergleichsweise langsame private Internetanschlüsse der Lehrenden überflüssig macht. Gleichzeitig sollen Austausch und Bereitstellung von VM-Images durch Verwendung eines Objektspeichers via S3-API flexibler und eine bessere Durchlässigkeit zur Cloud zum direkten Austausch erreicht werden. Dies wird im Zuge des PePP-Verbundprojekts in den kommenden zwei Jahren umgesetzt und an weiteren Universitäten im Land evaluiert.

8 Fazit

bwLehrpool stellt sich als Plattform zum effizienten Betrieb von PC-Pools an Hochschulen und Universitäten dar, die wartungsarm und flexibel eine Reihe von Standardaufgaben in der Lehre an Hochschulen abdeckt und diese bei der Umsetzung moderner IT-Strategien unterstützt. Es ist ein sehr robustes System entstanden, welches seit fünf Jahren als allgemeiner Landesdienst und seit einiger Zeit darüber hinaus angeboten wird, und knapp 6000 Pool-PCs[9] und weitere Maschinen steuert. Kooperative

[8] https://hochschulforumdigitalisierung.de/de/blog/online-proctoring
[9] Stand April 2022

Ansätze helfen bei einer Konzentration auf die Grundversorgung und breit nachgefragte IT-Dienste, da diese von „Economies of Scale“ profitieren. Sie eröffnen neue Chancen bei der Ausdifferenzierung IT-gestützter Lehre und stärken glaubwürdig die Digitale Souveränität der eigenen Einrichtung. Auf diese Weise werden zentrale Strategieelemente wie Open Source Orientierung und Chancengleichheit in der Bildung adressiert. Dieses schafft einen gewichtigen Gegenpol zu den vielen kommerziellen IT-Anbietern, die mit Macht in den lukrativen Bildungsmarkt drängen. Die doppelte Zustandslosigkeit des Systems eröffnet ein weiteres Einsatzgebiet, die Nutzung als resiliente Desaster-Recovery-Plattform [Su21].

Teile der dargestellten Entwicklungen wurden durch das Projekt bwGPUL erbracht, welches durch Fördermittel des Ministeriums für Wissenschaft, Forschung und Kunst Baden-Württemberg ermöglicht wurden. Das Verbundprojekt PePP (Partnerschaft für innovative E-Prüfungen im Projektverbund der baden-württembergischen Universitäten) wird durch die “Stiftung Innovation in der Hochschullehre” gefördert.

9 Literatur

[Ba19] Bauer, Jonathan; Messner, Manuel; Janczyk, Michael; Suchodoletz, Dirk von; Wiebelt, Bernd; Rasche, Helena: A Sorting Hat For Clusters. Dynamic Provisioning of Compute Nodes for Colocated Large Scale Computational Research Infrastructures, Universität Tübingen, 2019

[CB17] Corbi, Alberto; Burgos, Daniel: Open Distribution of Virtual Containers as a Key Framework for Open Educational Resources and STEAM Subjects. In: Electronic Journal of e-Learning Bd. 15 (2017), Nr. 2, S. 126-136

[KD21] Kätzel, Charlotte; Durst, Carolin: Mit agilen Blended Learning-Methoden erfolgreich durch die Pandemie: Was Unternehmen von den Hochschulen lernen können. In: HMD Praxis der Wirtschaftsinformatik Bd. 58 (2021), Nr. 4, S. 842–857

[RSB19] Rettberg, Simon; Suchodoletz, Dirk von; Bauer, Jonathan: Feeding the Masses: DNBD3. Simple, efficient, redundant block device for large scale HPC, Cloud and PC pool installations, Universität Tübingen, 2019

[Ri16] Ritter, Steffen; Trahasch, Stephan; Slotosch, Sven; Suchodoletz, Dirk von; Münchenberg, Jan: bwLehrpool: Durchführung von elektronischen Prüfungen in virtualisierten Umgebungen. In: DeLFI 2016–Die 14. E-Learning Fachtagung Informatik, Gesellschaft für Informatik eV, 2016

[Su21] Suchodoletz, Dirk von; Leendertse, Jan; Wiebelt, Bernd; Messner, Manuel Johannes; Janczyk, Michael: Stateless System Remote-Boot Als Business-Continuity-Konzept. In: Sicherheit in Vernetzten Systemen 28. DFN-Konferenz. Hamburg, 2021 — ISBN 978-3-7528-9805-7, S. F1–F21

Barriers to Digital Higher Education Teaching during the COVID-19 Pandemic from Teachers' and Students' Perspectives

Nicole Draxler-Weber [1], Henning Brink [2] and Sven Packmohr [3]

Abstract: During the COVID-19 pandemic, the rapid transition to digital teaching enabled higher education institutions (HEIs) to continue teaching. The strict execution exposed barriers that both teachers and students have faced towards digital higher education teaching around the world. In this paper, the barriers from both perspectives are identified and systematically processed. For this purpose, a quantitative survey of 396 students from HEIs in Sweden, Türkiye, and Germany was conducted. The students' barriers were identified and assigned to categories based on teachers' barrier categories, which were analyzed in a pre-study by conducting a literature review. The teachers' barrier categories could be confirmed by the students' survey. However, within the subcategories, the two perspectives differ. All categories and subcategories are described in detail so that this contribution offers an overview of barriers that have to be overcome if digital higher education teaching will be followed in the future.

Keywords: Digital teaching, barriers, teachers, students, higher education institutions, COVID-19 pandemic

1 Introduction and Theoretical Background

Even before the COVID-19 pandemic, there were tensions in higher education institutions (HEIs) on how to support, engage, and graduate underserved students. [PF20]. These tensions were enforced even more to all student groups, as HEIs experienced significant challenges due to the COVID-19 pandemic. Presence-based teaching formats had to be stopped and solely digital teaching had to be adopted within a quite short time. Teaching in its original form had disappeared, leading to a transition for both teachers and students. The term "teacher" is used to describe the individuals who instruct classes in HEIs. These range from tutors, or student assistants to professors. The term "student" refers to learners

[1] Osnabrück University, Department Business Studies/Organization and Information Systems, Katharinenstraße 1, 49069 Osnabrück, nicole.draxler-weber@uni-osnabrueck.de,
https://orcid.org/0000-0002-4462-227X
[2] Osnabrück University, Department Business Studies/Organization and Information Systems, Katharinenstraße 1, 49069 Osnabrück, henning.brink@uni-osnabrueck.de,
https://orcid.org/0000-0003-4426-1790
[3] Malmö University, Department of Computer Science and Media Technology, Nordenskiöldsgatan 1, 21119 Malmö, sven.packmohr@mau.se,
https://orcid.org/0000-0002-3488-708X

doi:10.18420/inf2022_123

at HEIs with completed school education. Although the digitization of instruction in higher education (HE) has been debated already before the pandemic, various elements prevented its implementation [Te20]. These elements are referred to as *barriers* during the course of this paper. The implementation of digital HE teaching was forced by the pandemic, which provided the opportunity for extensive research on digital teaching barriers that affected both teachers and students. Due to the rising tensions mentioned above, it is necessary to evaluate barriers from teachers' and students' perspectives to understand both perspectives. Even if barriers occur, the way both groups perceive them might help in overcoming these barriers. Furthermore, many research studies predominantly focused on the approaches adopted by certain HEIs or nations [PGK20], as well as case studies [MS20]. There is little to no international research where the above-mentioned barriers of both groups from different countries were systematically analyzed on an international level with a high number of participants and a merger of results from both stakeholder groups. HEIs around the world should benefit from pandemic experiences to improve their digital teaching in the future. Therefore, the experiences of both groups are of interest for the further development of HEIs.

Consequently, this research addresses the following research question (RQ):

RQ: What are the barriers experienced by teachers and students in adapting to digital higher education teaching during the COVID-19 pandemic?

To answer the RQ a two-stage research process is adopted, starting with an already published literature review on the teachers' barriers and an analysis of students' experiences collected via a questionnaire. We have chosen this two-step approach to combine insight from literature with empirical data. Also, the literature provides insight into teachers' barriers, already. Thus, we add students' perspective to get an overview of both perspectives. In section 2, the *Method* is described in detail. Thereafter, the results are presented in section 3 and discussed in section 4. The paper is finalized with *Conclusions and Implications* in section 5, where limitations are indicated, and an outlook is discussed.

2 Method

2.1 Literature Review

To resolve the RQ, our present study takes the results of a pre-study based on a literature review further [DVB22]. The pre-study focused on barriers to digital HE teaching from a teacher's perspective and was conducted in four steps based on the methodology recommended by Schryen [Sc15]. After defining the relevant search terms to detect barriers to digital teaching during COVID-19 in HEIs, 369 articles were identified in the databases Google Scholar, AISnet, and ERIC (Education Resources Information Center) in the first step. After, duplicates were eliminated, and the content of the articles was

checked if they targeted the topic and were published in scientific journals or conference proceedings. This second step resulted in 109 articles. In a third step, the fit was rechecked in-depth by examining the content to ensure the focus on teachers and include articles with sound data collection. Field and practice reports without a clear scientific methodology and literature reviews were eliminated. Finally, 13 papers were identified as relevant studies in step 4 based on the systematic literature review [DVB22].

Then, the barriers mentioned in these 13 articles were identified and various categories were determined based on these barriers. In the course of the development of inductive categories, various findings were first allocated to subcategories, which were then split into categories [Pu19]. The result of the pre-study is a concept matrix regarding teachers' barriers to digital HE teaching containing eight different categories with two to three subcategories [DVB22].

2.2 Questionnaire

Since the literature review focused on barriers from teachers' perspective, we conducted an online survey form to acquire the student perspective. This allowed us to identify similarities and differences, resulting in a more holistic picture.

To do so, we aimed to gather a diversified sample. We surveyed students from three different HEIs in Sweden, Türkiye, and Germany. This allowed us to acquire a cross-country and thus an international insight. We applied a convenience sampling strategy [EMA16] including students from bachelor's and master's programs of different semesters. The sample contains both the view of students who studied at the respective HEI before the COVID-19 pandemic and students who started their studies during the pandemic. All the students questioned had been participants in elective courses in business studies programs with content linked to information systems. These courses were digitally transformed and conducted entirely remotely due to contact restrictions and disease prevention interventions during the COVID-19 pandemic. All participants responded to the same web-based survey voluntarily and anonymously. In sum 396 students responded and completed the survey between January 2021 and January 2022. The sample consists of roughly balanced proportions of male (40%) and female (56%) participants. No participant stated to be diverse, while 4% of participants did not specify their gender. Further, 84% of the participants were between 20 and 30 years old.

In the survey, all the students were solicited to state the three most substantial barriers to digital teaching in open-ended questions: "In your opinion, what are the 3 strongest barriers to digital teaching?" After collecting the data, it was analyzed in a second step. Data from 1190 statements were particularly determined about significant barriers based on the participant responses. After the collection, the statements were coded based on the categories and subcategories of the pre-study. Thus, a deductive coding procedure was adopted [Az18]. However, 808 statements did not fit the pre-study concept matrix based on the literature review. Individually, the three authors of this paper coded these statements inductively [Az18]. The inductively coded statements were discussed and clustered into subcategories in several rounds of revisions to ensure inter-coder agreement between all

authors. The result is an adapted barrier overview, including both students' and teachers' perspectives, described in section 3 and displayed in Tab. 1.

3 Findings

All identified teachers' barriers from the pre-study (literature review) and all students' barriers identified in the survey are submitted in Tab. 1. The identified barriers could be assigned to eight categories: *Technical Resources, Interaction, Skills, Didactics, Workload, Health, Personal Readiness*, and *Framework Conditions*. Each category is allocated at least two subcategories. In the further course, the categories with their subcategories are described in more detail. For this purpose, quotations from the questionnaire are marked with "[QS]", whereas findings from the literature review are marked according to their respective publication source.

Technical Resources. *Technical Resources* are essential for conducting and taking part in digital HE teaching. Therefore, both teachers and students perceive this as a barrier when there is a *lack of technical resources* and/or an *insufficient internet connection*. A *lack of technical resources* occurs, for example, when equipment is missing [GC20] or required software [MS20] and networks are only accessible to a limited extent [Al21]. Such a lack can appear either on the side of the teachers or the students. In the analyzed survey, "the economic status of some students" [QS] was mentioned as a reason for the deficiency on the students' side. The subcategory *insufficient internet connection* can include general connection problems [HX20, KT21, MS20, Pe20], lack of internet bandwidth [KT21], or even the absence of internet access [Ku21, KT21] which were also all specified by the students in the survey.

Interaction. Teachers and students see the *Interaction* in digital teaching during the COVID-19 pandemic seriously disrupted. This has been expressed in different ways, which is evident in the various subcategories: *lack of social interaction*, *lack of discussion and participation, invisible reactions*, *miscommunication*, and the *low possibility to ask and get help*. Teachers and students miss social contact and determine a *lack of social interaction*. According to the students, "the connection with students and teachers has become inferior" [QS]. Personal relationships with teachers and fellow students cannot be established. Besides, the teachers lack social contact with their colleagues [Al20, EC20, MS20, Pe20, PGK20]. The interaction while teaching is also considered insufficient, as expressed by the subcategory *lack of discussion and participation*. Teachers miss the interaction with students while teaching and students confirm online discussions are difficult to come up with because the necessity "to unmute to speak leads to students not talking spontaneously" [QS]. Moreover, the students rate their active participation as not adequate. Digital teaching "creates a resistance to discussing because people are still uncomfortable" [QS]. In addition, *invisible reactions* are seen as a barrier. Teachers and students cannot see each other's reactions during digital teaching. Teachers cannot know if students can and do follow their teaching [KT21] because nonverbal cues are lost [Pe20].

This was also confirmed by the students in the results of the survey: “It’s more difficult to be understood without having eye contact, it’s difficult to explain thing[s] [be]cause you’ll never know if the person on the other side of the screen is looking at you or something else” [QS]. This often ends in misunderstandings and misinterpretations that otherwise would not have arisen. For this reason, general *miscommunication* was also mentioned as a barrier by the students. In this context, students criticize not only a misinterpretation of the teacher’s statements but also of the students among themselves. Another addition thanks to the student survey is the *low possibility to ask and get help*. Within digital teaching, there are fewer opportunities to ask questions or seek help. Students see it as difficult to get the attention they need online to ask questions or get individual help. In particular, the possibility “to stay after class and get help sometimes” [QS] is missing.

Category	Subcategory	Teachers	Students
Technical Resources	Lack of technical resources	X	X
	Insufficient internet connection	X	X
Interaction	Lack of social interaction	X	X
	Lack of discussion and participation	X	X
	Invisible reactions	X	X
	Miscommunication		X
	Low possibility to ask and get help		X
Skills	Lack of digital competencies	X	X
	Lack of training offers	X	
	Lack of self-management		X
Didactics	Limited possibilities for lecture design	X	X
	Lack of practical exercise	X	X
	Lack of skill transfer	X	X
	Lack of knowledge transfer		X
Workload	Higher workload	X	X
	Lack of time for research	X	
	Laborious group work		X
Health	Higher stress level	X	X
	Isolation	X	X
	Bodily discomforts		X
	Concentration problems		X
Personal Readiness	Lack of flexibility	X	X
	Lack of trust	X	X
	Lack of motivation		X
	Fear of change	X	X
Framework Conditions	Restrictive legal frameworks	X	
	Insufficient institutional frameworks	X	X
	Insufficient learning environment		X

Tab. 1: Teachers’ and students’ barriers to digital higher education teaching

Skills. Digital HE teaching demands certain skills from both teachers and students. A *lack of digital competencies* was observed in both groups. Among teachers, it is characterized by their inability to manage technologies [KT21] and to use online tools and e-learning systems [Al21, HX20]. However, some students also lack the digital competencies they need to learn digitally. That lack is considered a reason for the disruption of teaching sessions: "Having to constantly check or guide someone can take the focus from the lecture and ultimately, students' learning possibilities" [QS]. Teachers see a *lack of training offers* as a further barrier [Ku21], whereby their digital competencies cannot be developed and raised [Al20]. In addition, students note a *lack of self-management*, which is strongly required in the context of digital teaching. Teachers cannot be "unstructured and unprepared" [QS], while students must plan their learning well, even if they "are not used to it" [QS].

Didactics. Digital HE teaching requires its own *Didactics.* Existing didactic concepts from conventional teaching cannot simply be transferred but must be adapted. This is seen as a barrier by teachers and students due to *limited possibilities for lecture design*, *lack of practical exercise*, *lack of skill transfer,* and *lack of knowledge transfer*. Digital teaching only enables *limited possibilities for lecture design* because active and collaborative teaching methods are partially possible [Al20]. In this regard, students observed "problems in the transfer of course contents" [QS]. Besides, a *lack of practical exercise* during the pandemic was determined. This mainly concerns practical skills that only can be thought in specific laboratories or workshops [Al21, EC20, GC20, Ku21, Ob20]. This also was confirmed by students who missed the practical learning: "Not being able to physically see and try things that would help the learning process" [QS]. Apart from the practical experience, the possibilities of skill transfer, especially soft skills [GC20], were also missing. Relating to the *lack of skill transfer,* students criticize the less practice of presentation skills. Even though presentations were held online, there are not seen "as reality-based as in physical classrooms" [QS]. Furthermore, the students pointed out the *lack of knowledge transfer*. Some students have difficulty keeping up in digital courses by "not fully understanding" [QS] the content or assignments.

Workload. The *Workload* of teachers has changed during the COVID-19 pandemic [Al20, Al21, PGK20] as has that of students. Digital HE causes a *higher workload* for teachers by the increased time to prepare for teaching online [EC20, HX20, KT21]. Due to the reason that digital teaching is also a novelty for the students, many questions arise for them and teachers spent a lot of time replying to messages [EC20, Wa21]. This leaves no time to pursue teachers' research projects [Wa21] *w*hich is summarized in the subcategory *lack of time for research.* Students perceive the increased workload of teachers by observing that their requests are not always answered by teachers. At the same time, the subcategory also represents a *higher workload* for students. Because of digital teaching, students observe an increase in assignments that need to be completed. Frequently, assignments must be submitted as group work, which has been very tedious for students. The *laborious group work* is marked by the fact that "without meeting physically the productivity, team spirit and quality often worsen" [QS]. Overall, group work is harder to organize and takes more of the student's workload.

Health. The *Health* of teachers and students is negatively affected by digital HE teaching during the pandemic. Both parties experienced a *higher stress level.* Teachers' increased stress level could result from external pressure to teach online [MS20], while students feel "stressed before and during lectures, due to a very tight schedule" [QS]. The home office regulations during the COVID-19 pandemic forced not only students to stay absent from campus but also teachers to work from home [MS20]. This led to *isolation* for both groups. Teachers lack a physical workplace where they meet colleagues, and students lack a campus where they meet other students. Not having to leave the house thus "sitting all the time" [QS] has also led to *bodily discomforts* in the view of students, like "problem with eyes" [QS]. Moreover, students observe *concentration problems*. Students find it much harder to maintain their concentration online during the whole lecture.

Personal Readiness. With the pandemic, digital teaching was launched very rapidly, so that neither teachers nor students could adjust to it. This is addressed within the barrier *Personal Readiness,* which includes the subcategories: *lack of flexibility*, *lack of trust*, *lack of motivation,* and *fear of change*. *Lack of flexibility* refers to conservatism among teachers [Al20] and students. One statement in the survey referred to "acquired habits" among all teaching participants, which first had to be broken. However, this proves to be difficult the larger the *lack of trust* towards digital teaching among teachers and students. Poor confidence in their own ability to use online technologies, or a general reluctance to adopt them, could further result in a *lack of trust* [KT21, Wa21]. Students also complain in the survey that teachers do not trust them and believe "that students care less about the lecture" [QS]. At the same time, students frequently referred to the *lack of motivation*. Students partly do not have the motivation "to get up in the morning and watch the digital class" [QS]. This is mainly owed to the fact that "a lot of people are very annoyed about having to study online and are therefore less likely to make an effort" [QS]. Furthermore, a *fear of change* can also be a barrier. Such fear can arise when technology is not considered advantageous [Wa21]. This barrier was also identified by the students but was only assigned to the group of teachers as "teachers fear" [QS].

Framework Conditions. External conditions can also be a barrier to digital HE teaching. These have been grouped in the *Framework Conditions* and include the subcategories *restrictive legal frameworks*, *insufficient institutional frameworks,* and *insufficient learning environment*. Teachers observe various requirements of digital teaching formulated by the government that have to be followed [Al21]. This barrier is summarized under *restrictive legal frameworks*. Besides, institutional framework conditions also need to be adapted to the new circumstances by creating curricula or consistent regulations in the HEIs [Al21, Ku21]. However, *insufficient institutional frameworks* are identified by both teachers and learners. In this context, students note inadequate examination conditions during digital teaching. Cheating in exams is much easier and should be better controlled. Students perceive the insufficient information from the HEIs as a barrier and demand more explanations, for example, about courses or exams. In addition, students criticize their *insufficient learning environment*. The pandemic forced all students to stay home, so learning no longer took place on campus. However, the learning environment at

home was not optimal for all students, for example, due to distractions like background noises, younger siblings, or a crowded house.

4 Discussion

From the results of section 3 regarding the barriers experienced by the teachers and students on the adaption of digital HE teaching during the COVID-19 pandemic, it is noticeable that the eight generic categories of teachers' barriers were also confirmed by students. Only in the specification of the individual categories through the subcategories, differences between teachers and students can be observed. The largest overlap is found within the category *Technical Resources*. Both teachers and students confirm that the *lack of technical resources* and an *insufficient internet connection* are essential barriers to digital HE teaching. For all other seven categories, differences can be identified at the subcategory level. Either existing subcategories did not find a denomination among students or additions had to be made to consider students' perspectives.

Subcategories that were not represented among the students' mentions are *lack of training offers*, *lack of time for research*, and *restrictive legal frameworks,* affecting three categories. Although students criticize the *lack of digital competencies* of teachers and students as well, they do not complain about the *lack of training offers*. Sánchez-Caballé et al. determine in a detailed systematic literature that students prefer to train their digital competencies in guided subjects and therefore less in additional training sessions [SGE20]. HEIs have fundamental two missions: teaching and research [Ma12]. Since students are not heavily involved in research projects besides their teaching modules, it is not surprising that the *lack of time for research* was not mentioned. This barrier is more specific to teachers in HEIs who have research assignments, besides their teaching responsibilities [Ma12]. The last subcategory that was not reflected among students was the *restrictive legal frameworks*. Even in the literature review of teachers' barriers, this subcategory has just single references. Especially in an analyzed paper from China, this subcategory was mentioned in the context of censorship. In China, online content can be easily blocked due to its sensitive content [HX20]. This can be summarized as a country-specific phenomenon, so it is less surprising that it was not mentioned among students in the present survey.

Except for the *Technical Resources* category, additions were made in all other seven categories. The following subcategories were added based on student statements: *miscommunication, low possibility to ask and get help, lack of self-management, lack of knowledge transfer, bodily discomforts, concentration problems, lack of motivation*, and *insufficient learning environment*. It is noticeable that most additions were determined within the *Interaction* subcategories and were strongly argued by the students. The addition of *miscommunication* clarifies in the digital teaching environment, students feel misunderstood or struggle with misconceptions about what teachers are saying. This is often related to *invisible reactions*, which were also criticized by the teachers. However,

the teachers do not focus on the misinformation this generated but rather on the difficulties of being able to follow the students' learning process [KT21]. In addition, students saw a *low possibility to ask and get help.* The teachers, on the other hand, saw a lot of guidance offered to students, which cost them a particularly high effort. This was mentioned in the *higher workload*, which the teachers attributed to the higher support requirements [EC20, Wa21]. So, it is less surprising that the *low possibility to ask and get help* was not seen as a barrier by teachers. But the compensation seems to have been insufficient in students' view. Otherwise, with the transition to digital teaching, students were required to become more independent in learning, because the learning plan was not monitored on campus anymore [MMM21]. This contrasts with the *lack of self-management* of students. This identified barrier reveals students were not ready to manage their teaching due to insufficient *Skills*. For the transition to digital teaching, preparations are necessary that includes also the students [MMM21]. In the context of *Didactics*, teachers assumed that especially the skill transfer [GC20] and generally the practical exercises were limited [Al21, EC20, GC20, Ku21, Ob20], which was also confirmed by students. However, students' statements also show even theoretical contents were not as well-received as in the presence. Although teachers saw *limited possibilities for lecture design*, it was not considered as harming to student knowledge transfer. This correlation to *lack of knowledge transfer* can now be drawn. The COVID-19 pandemic had led to several mental health issues due to lockdowns and quarantine policies [GJA20]. It is, therefore, no surprise that *isolation* and *higher stress level* affecting mental health were seen as barriers by teachers and students. Besides these aspects, students also noted *bodily discomforts* and *concentration problems*. This indicates that, in addition to mental health, physical health also suffered under the conditions of digital teaching during the pandemic. Furthermore, the pandemic-caused conditions were also a reason for the *lack of motivation* among students [PF20]. The isolation and the constraint of only being able to learn online emerged as demotivating factors. At the same time, the learning environment at home was insufficient and strengthened students' wish to be able to study on campus again. Other research has revealed that students are more sensitive to external factors compared to teachers and online teaching should also consider students' personal needs [LS21].

In sum, the results revealed both differences and similarities in students' and teachers' perceptions of barriers. While equal perceptions of teachers and students in digital teaching can create a sense of cohesion, differences could weaken the success of digital HE teaching and create negative effects. Even though both groups aim to provide and receive a valuable education. Existing research has already shown that teachers and students do not recognize all problems of the other group, e.g. concerning workload or motivation. [NK20], or accurate information about the other part, is not always present [BMS16]. The results of our study indicate a similar pattern. As mutual understanding is of great importance [SS20], our results can help to close gaps between the two groups.

5 Conclusions and Implications

The COVID-19 pandemic led to a radical shift into digital teaching, so the experiences from this period provide valuable insights into the barriers toward digital HE teaching faced by teachers and students. These experiences represent a good opportunity to examine the barriers in detail to be able to overcome them in establishing digital teaching in the future. This paper contributes to this by identifying the barriers to digital teaching during the pandemic from two perspectives. For this purpose, a pre-study by the authors is used in which the barriers of teachers were identified from the literature. A survey of students from different countries was built upon this by mapping the students' statements to the categories from the pre-study. If necessary, the categories with their subcategories were adjusted or supplemented. In this way, a holistic picture of teachers' and students' barriers to digital teaching during the pandemic consisting of the eight categories emerges: *Technical Resources, Interaction, Skills, Didactics, Workload, Health, Personal Readiness*, and *Framework Conditions*. Our results are valuable for institute directors and teachers. They can address barriers in advance and therefore pave the way for a successful implementation of digital HE teaching. Researchers, on the other hand, receive a solid carefully developed barrier overview that can be reviewed and expanded, especially considering a post-pandemic era.

Despite the conscientious work of the authors, this paper is not free of limitations. The conduct of the used pre-study already took place in the spring of 2021, in the early stages of the pandemic. At the time, only a few sources were available that dealt with the subject scientifically. Thus, the pre-study results are based on a literature review with 13 analyzed articles, according to a pre-defined search string. It can be assumed that, in the meantime, articles were published on the subject that were not considered and that would have led to a different teachers' barriers result. However, the categories are rather generic and specified in the subcategories. The predetermined categories from the preliminary study provided a strong direction in the evaluation, as it was a deductive category formation. Only statements by the students that could not be assigned to any of the pre-defined categories were added inductively. It can be assumed that an independent evaluation of the student data and an inductive category formation would have led to different subcategories. Furthermore, we applied a convenience sampling strategy, which is a type of non-probability sampling and is helpful to explore a field but is prone to bias [St17]. Future research should therefore cross-check the results with other sampling strategies. Nevertheless, this methodological approach has enabled an overview of the barriers of both groups. In this way, it becomes clear what barriers to digital teaching existed from both sides during the pandemic. This provides a basis for further research, which should investigate how these barriers can be overcome and formulate recommendations for action. Further quantitative studies could provide insights into the frequency and intensity of perceived barriers. Overall, the experience gained during the pandemic should be used to remove barriers to digital teaching and to advance digitization in HE teaching.

Bibliography

[Al20] Almazova, N. et al.: Challenges and opportunities for Russian higher education amid COVID-19: Teachers' perspective. In: Education Sciences Bd. 10, Multidisciplinary Digital Publishing Institute, Nr. 12, p. 368, 2020.

[Al21] Aldulaimi, S. H.: E-Learning in Higher Education and Covid-19 Outbreak: Challenges and Opportunities. In: Psychology and Education Journal Bd. 58, Nr. 2, pp. 38–43, 2021.

[Az18] Azungah, T.: Qualitative research: deductive and inductive approaches to data analysis. In: Qualitative Research Journal Bd. 18, Nr. 4, pp. 383–400, 2018.

[BMS16] Borghi, S.; Mainardes, E.; Silva, É.: Expectations of higher education students: a comparison between the perception of student and teachers. In: Tertiary Education and Management Bd. 22, Nr. 2, pp. 171–188, 2016.

[DVB22] Draxler-Weber, N.; Voigt, C.; Blömer, L.: Teachers' Barriers to digital Higher Education Teaching during the COVID-19 Pandemic. In: Wirtschaftsinformatik 2022 Proceedings. 4, 2022.

[EC20] Egan, K.; Crotty, Y.: Sustaining a Prolonged Pivot: Appraising Challenges Facing Higher Education Stakeholders in Switching to Online Learning. In: International Journal for Transformative Research Bd. 7, ERIC, Nr. 1, pp. 1–9, 2020.

[EMA16] Etikan, I.; Musa, S. A.; Alkassim, R. S.: Comparison of Convenience Sampling and Purposive Sampling. In: American Journal of Theoretical and Applied Statistics. Vol. 5, No. 1, pp. 1-4, 2016.

[GC20] Gonçalves, E.; Capucha, L.: Student-centered and ICT-enabled learning models in veterinarian programs: what changed with COVID-19? In: Education Sciences Bd. 10, Multidisciplinary Digital Publishing Institute, Nr. 11, p. 343, 2020.

[GJA20] Gunawan, J.; Juthamanee, S.; Aungsuroch, Y.: Current Mental Health Issues in the Era of Covid-19. In: Asian Journal of Psychiatry Bd. 51, p. 102103, 2020.

[HX20] He, W.; Xiao, J.: The Emergency Online Classes during COVID-19 Pandemic: A Chinese University Case Study. In: Asian Journal of Distance Education Bd. 15, ERIC, Nr. 2, pp. 21–36, 2020.

[Ku21] Kummitha, H. R. et al.: Coronavirus disease 2019 and its effect on teaching and learning process in the higher educational institutions. In: Higher Education for the Future Bd. 8, SAGE Publications Sage India: New Delhi, India, Nr. 1, pp. 90–107, 2021.

[KT21] Kono, K. G.; Taylor, S.: Using an Ethos of Care to Bridge the Digital Divide: Exploring Faculty Narratives During a Global Pandemic. In: Online Learning, 2021.

[LS21] Lei, S. I.; So, A. S. I.: Online Teaching and Learning Experiences During the COVID-19 Pandemic – A Comparison of Teacher and Student Perceptions. In: Journal of Hospitality & Tourism Education Bd. 33, Routledge, Nr. 3, pp. 148–162, 2021.

[MMM21] Maphalala, M. C.; Mkhasibe, R. G.; Mncube, D. W.: Online Learning as a Catalyst for Self-directed Learning in Universities during the COVID-19 Pandemic. In: Research in Social Sciences and Technology Bd. 6, Nr. 2, pp. 233–248, 2021.

[Ma12] Martin, B. R.: Are universities and university research under threat? Towards an evolutionary model of university speciation. In: Cambridge Journal of Economics Bd. 36, Nr. 3, pp. 543–565, 2012.

[MS20] Marshalsey, L.; Sclater, M.: Together but Apart: Creating and Supporting Online Learning Communities in an Era of Distributed Studio Education. In: International Journal of Art & Design Education Bd. 39, Wiley Online Library, Nr. 4, pp. 826–840, 2020.

[NK20] Niemi, H. M.; Kousa, P.: A Case Study of Students' and Teachers' Perceptions in a Finnish High School during the COVID Pandemic. In: International Journal of Technology in Education and Science Bd. 4, Nr. 4, pp. 352–369, 2020.

[Ob20] O'Brien, W. et al.: Implications for european physical education teacher education during the COVID-19 pandemic: a cross-institutional SWOT analysis. In: European Journal of Teacher Education Bd. 43, Taylor & Francis, Nr. 4, pp. 503–522, 2020.

[PF20] Patel, V.; Field, K.: Retaining Underserved Students: Strategies for Success in a Post-pandemic Era. Washington, D.C., Chronicle of Higher Education, 2020.

[Pe20] Perron, N. J. et al.: Online synchronous clinical communication training during the covid-19 pandemic. In: Advances in medical education and practice Bd. 11, Dove Press, p. 1029, 2020.

[PGK20] Popova, O.; Gagarina, N.M.; Karkh, D. A: Digitalization of Educational Processes in Universities: Achievements and Problems. In: Advances in Social Science, Education and Humanities Research Bd. 437, p. 738, 2020.

[Pu19] Puppis, M.: Analyzing talk and text I: Qualitative content analysis. In: The Palgrave handbook of methods for media policy research : Springer, pp. 367–384, 2019.

[SGE20] Sánchez-Caballé, A.; Gisbert-Cervera, M.; Esteve-Mon, F.: The digital competence of university students: a systematic literature review. In: Aloma: Revista de Psicologia, Ciències de l'Educació i de l'Esport Bd. 38, Nr. 1, pp. 63–74, 2020.

[Sc15] Schryen, G.: Writing Qualitative IS Literature Reviews—Guidelines for Synthesis, Interpretation, and Guidance of Research. In: Communications of the Association for Information Systems Bd. 37, Nr. 1, 2015.

[SS20] Silva, O.; Sousa, Á.: PERCEPTION OF TEACHERS AND STUDENTS ABOUT TEACHING AND LEARNING IN THE PERIOD OF COVID-19 PANDEMIC. In: Online Conference, pp. 4832–4838, 2020.

[St17] Stern, M. J.; Bilgen, I.; McClain, C.; Hunscher, B.: Effective Sampling from Social Media Sites and Search Engines for Web Surveys: Demographic and Data Quality Differences in Surveys of Google and Facebook Users. In: Social Science Computer Review Bd. 35 (2017), Nr. 6, S. 713–732.

[Te20] Tesar, M.: Towards a Post-Covid-19 'New Normality?': Physical and Social Distancing, the Move to Online and Higher Education. In: Policy Futures in Education Bd. 18, Nr. 5, pp. 556–559, 2020.

[Wa21] Watermeyer, R. et al.: COVID-19 and digital disruption in UK universities: Afflictions and affordances of emergency online migration. In: Higher Education Bd. 81, Springer, pp. 623–641, 2021.

Künstliche Intelligenz in der Umweltinformatik (KIU-2022)

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1443

3. Workshop Künstliche Intelligenz in der Umweltinformatik

Andreas Abecker [1], Julian Bruns [2], Stefan Naumann [3]

Abstract: In den letzten Jahren werden die vom Menschen verursachten Umweltprobleme, wie Klimawandel, Verlust der Artenvielfalt, Verschmutzung der Meere usw. immer offenkundiger, in ihren negativen Auswirkungen für den Menschen immer dramatischer und ihr besseres Verständnis als Grundlage für kraftvolle Lösungsansätze wird immer bedeutender und drängender, wenn wir unsere gegenwärtige Lebensweise mit einem hohen Maß an materiellem Wohlstand und politisch-gesellschaftlicher Stabilität auch nur einigermaßen wahren wollen. Die Informatik und auch die Künstliche Intelligenz können und sollten eine hervorgehobene Rolle einehmen, wenn es darum geht, komplexe natur-, umwelt-, erd- und lebenswissenschaftliche Zusammenhänge besser zu erfassen, zu überwachen, zu verstehen und aus diesem Verständnis heraus effizientere und effektivere technische Lösungsansätze zum Umgang mit den großen Umweltproblemen zu entwickeln. In der KIU-Workshopreihe werden in diesem großen Themenkomplex neue innovative Ideen vorgestellt, offene Forschungsfragen und Lösungsansätze diskutiert und Anwendungsbeispiele aus den Schnittfeldern von KI und Umweltanwendungen präsentiert.

Keywords: Umweltinformatik, Künstliche Intelligenz, Maschinelles Lernen, Umweltmonitoring, Umweltschutz, Nachhaltigkeit

1 Ziele und Motivation des Workshops „Künstliche Intelligenz und Umweltinformatik“

Die Umweltinformatik befasst sich interdisziplinär mit der Analyse und Bewertung von Umweltsachverhalten. Aus IKT-Sicht spielen dabei beispielsweise Simulationen komplexer Systeme, Geographische Informationssysteme und räumliche Datenanalyse, Messnetze und Sensordatenverarbeitung sowie Fernerkundung und Bildverarbeitung eine große Rolle. Die Umweltinformatik findet ihre wichtigsten Anwendungen traditionell in der Wissenschaft und in der öffentlichen Verwaltung (Natur- und Umweltschutz, Umweltdatenportale, Katastrophenschutz, Verbraucherschutz, Wassermanagement etc.), aber in jüngerer Zeit zunehmend auch in Wirtschaft und Gesellschaft, um dort ressourcenschonendere Lebens-, Arbeits- und Produktionsweisen zu schaffen. Enge Bezüge und Überlappungen zur Umweltinformatik ergeben sich auch aus der Agrarinformatik, der Hydroinformatik, der Energieinformatik, der Gesundheitsinformatik sowie im Bereich Green IT. In jüngerer Zeit entsteht ein eng verwandtes interdisziplinäres Themenfeld unter Schlagworten wie „ICT for Sustainability (ICT4S)“ oder „Computational Sustainability“.

[1] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe andreas.abecker@disy.net
[2] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe julian.bruns@disy.net
[3] Hochschule Trier, Umwelt-Campus Birkenfeld, Institut für Softwaresysteme, Postfach 1380, 55761 Birkenfeld s.naumann@umwelt-campus.de

doi:10.18420/inf2022_124

Die Umweltinformatik betrachtet zumeist komplexe ökologische Prozesse, die (noch) nicht vollständig bekannt, verstanden oder verstehbar sind (chaotische Systeme), deren Verhalten nur approximativ oder vereinfachend modelliert, simuliert oder vorhergesagt werden kann oder soll (z.B. aus Aufwandsgründen) und bei deren Beobachtung zuweilen wichtige Größen unbekannt sind oder nur geschätzt werden können. Zusätzlich müssen bei der Betrachtung von Ökosystemen noch weitere Systeme, die aus Informatiksicht nicht einfach zu behandeln sind, wie z.B. das Wetter, betrachtet werden, da sie direkte oder indirekte Einflüsse auf die betrachteten Systeme ausüben. Aus Sicht der Anwendung müssen häufig Entscheidungen unter Unsicherheit getroffen und schwierige Abwägungen in Zielkonflikten getroffen werden, die Fach- und Erfahrungswissen erfordern (z.B. in Planungsverfahren, im Notfallmanagement, bei der Politikgestaltung). Die entsprechenden Fragestellungen haben vielfältige Wechselwirkungen zu schwierigen Themen wie z.B. Klimawandel, Energiewende, Biodiversität und Nachhaltigkeit, aber auch mit großen Technologietrends wie Smart Cities, Smart Agriculture, Smart Mobility oder Smart Grids.

Die Anwendung von Methoden und Technologien der KI drängt sich also auf. Im Rahmen der INFORMATIK 2020 fand der Workshops KIU das erste Mal statt, mit der Idee, eine neue Community zum Themenfeld „KI und Umweltinformatik“ zu begründen. Teilnahme und Verlauf von KIU-2020 waren ermutigend. Es gab einen eingeladenen und 8 eingereichte und begutachtete Vorträge. Bei der Online-Durchführung waren permanent etwa 30 Zuhörer:innen präsent, in Spitzenzeiten waren es sogar bis zu 50 – bei insgesamt über 100 unterschiedlichen Tagungsteilnehmern im Workshop. Auch bei der zweiten Durchführung KIU-2021 gab es einen eingeladenen sowie 8 begutachtete Fachbeiträge und gute Teilnehmerzahlen. Wir hoffen, dass KIU-2022 erstmals (auch) als Präsenzveranstaltung durchgeführt werden kann und die Tiefe der fachlichen Diskussionen durch den persönlichen Austausch noch gewinnt.

2 Thematische Schwerpunkte des Call-for-Papers

KI-Technologien für die Umwelt (nicht ausschließliche Liste):

- Big / Smart / Linked / Open Data
- Bildverarbeitung und Fernerkundung
- Data Mining, Machine Learning, Deep Learning
- Digital Twins
- ELSI-Aspekte zur Umwelt-KI, Responsible AI
- Explainable AI, Human-Centered AI
- Multiagentensysteme
- Multimodale Interfaces, AR, VR und KI

- Robotik
- Semantische Technologien
- Soft Computing / Computational Intelligence
- Spatial Data Mining, Location Intelligence
- Sprachverarbeitung
- Unsicherheit und Vagheit
- Wissensbasierte Systeme

Umweltanwendungen mit KI-Unterstützung (nicht ausschließliche Liste):

- Betriebliche und behördliche Umweltinformationssysteme, Umweltmonitoring
- Biodiversität
- Computational Sustainability
- Erneuerbare Energien und Energiewende
- Green IT und Energiemanagement
- Katastrophenschutz und -management (aus Umweltsicht)
- Klimawandel
- Nachhaltige Produktion, Kreislaufwirtschaft
- Ressourcenschutz und Landmanagement
- Smart Agriculture, Smart Forestry (aus Umweltsicht)
- Smart City, Smart Mobility (Umweltaspekte)
- Umweltbildung
- Verbraucherschutz (Umweltaspekte)
- Wasser 4.0

3 Eingereichte Fachbeiträge

Auch bei der dritten Ausführung des Workshops konnten erfreulicherweise sowohl die betrachteten Umweltthemen als auch die verwendeten KI-Technologien nochmals erweitert werden gegenüber den Vorträgen der ersten beiden Workshop-Durchführungen. Während 2020 und 2021 der thematische Schwerpunkt stark auf dem Wassermanagement und technisch stark auf Deep Learning lag, bieten die 6 Fachvorträge im Jahr 2022 als zusätzliche

Fachthemen bspw. die Steuerung von Kläranlagen, das Smart Farming und die Kreislaufwirtschaft und als zusätzliche KI-Technologie das fallbasierte Schließen. Auch pragmatische und praxisorientierte Fragestellungen, um durch die KI-Anwendung tatsächlich kurzfristige Nutzeneffekte für die Umwelt zu erzielen, finden größere Beachtung als früher. Das ist vielversprechend für eine stetige Ausweitung der KIU-Community und lässt auf einen kurzweiligen und informativen Workshop hoffen. Auch Themen aus früheren Workshops wie Wassermanagement, Geodatenanalyse und Geodatenstandards, Architekturen zur Einbettung und Verwendung von KI-Verfahren sowie Deep Learning sind wieder vertreten.

4 Programmkommitee

- Dr. Ansgar Bernardi; Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern; https://www.dfki.de/
- Dr. Matthias Budde; Disy Informationssysteme GmbH, Karlsruhe; http://www.disy.net/
- Prof. Dr. Frank Fuchs-Kittowski; Hochschule für Technik und Wirtschaft, Berlin; https://www.htw-berlin.de/
- Dr. Katharina Glock; FZI Forschungszentrum Informatik, Karlsruhe; https://www.fzi.de/
- Dr. Desirée Hilbring; Fraunhofer IOSB, Karlsruhe; https://www.iosb.fraunhofer.de/
- Dr. Julian Huber; FZI Forschungszentrum Informatik, Karlsruhe; https://www.fzi.de/
- Prof. Dr. Christian Jolk; Technische Hochschule Ostwestfalen-Lippe, Höxter; https://www.th-owl.de/
- Dr. Sven Lautenbach; Universität Heidelberg, Heidelberg; https://www.geog.uni-heidelberg.de/
- Dr. Tanja Liesch; Karlsruhe Institut für Technologie, Karlsruhe; http://www.kit.edu/
- Prof. Dr. Jens Nimis; Hochschule Karlsruhe Technik und Wirtschaft, Karlsruhe; https://www.hs-karlsruhe.de/
- Dr. Steffen Thoma; FZI Forschungszentrum Informatik, Karlsruhe; https://www.fzi.de/
- Dr. Marc Wieland; Deutsches Zentrum für Luft- und Raumfahrt e.V. (DLR), Oberpfaffenhofen; https://www.dlr.de/eoc

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1447

Energieeinsparung auf Kläranlagen mit fallbasierter Steuerung (Cased Based Reasoning)

Günter Müller-Czygan (Hof), Jürgen Wiese (Magdeburg), Viktoriya Tarasyuk (Hof), Ralf Tschepetzki (Madgeburg)

Im Urlaubsort Heringhausen am Diemelsee (Nordhessen) leben ca. 400 Einwohner. In den Sommermonaten kommen aber bis zu 4.000 Übernachtungsgäste und am Wochenende weitere 1.000 Tagesgäste. Die im Jahr 2020 neu errichtete Kläranlage Heringhausen stellt aufgrund der besonderen Zulaufanforderungen durch den Tourismus das Herzstück der Datenanalyse und modernen Prozesssteuerung für die gesamte Abwassersammlung, -ableitung und -reinigung am Diemelsee dar. Hier werden zukünftig zudem alle wasserwirtschaftlichen Einrichtungen der Gemeinde datentechnisch zusammenlaufen, sodass die Kläranlage die Steuerungszentrale für eine zukunftsfähige, vernetzte Abwasserentsorgung der Gemeinde Diemelsee darstellt. Um die unterschiedlichen und teilweise sehr gegensätzlichen Zulaufbedingungen in einer integrierten Steuerungslogik berücksichtigen zu können, sollte erstmals in der deutschen Wasserwirtschaft ein erweiterter Analysealgorithmus nach dem CBR-Verfahren (Cased Based Reasoning) als Entscheidungshilfesystem eingesetzt werden. Übliche abwassertechnische Steuerungen arbeiten eher als Regelsystem, d.h. Anpassungen wichtiger Prozessparameter wie z.B. der sehr energieintensiven Drucklufterzeugung zur Versorgung von Abwasserbakterien mit Sauerstoff erfolgen erst, wenn Sensoren eine Veränderung am Prozessende erkennen. Zwar gibt es in der Abwasserwirtschaft seit vielen Jahren bereits Systeme, die anhand verschiedener Messparameter Veränderungen im Voraus erkennen können, diese sind aber noch immer sehr ungenau und meist auf einzelne Parameter beschränkt.

Der Mensch als Vorhersagesystem – Erfahrung schlägt Berechnung

Eine aktuelle Studie der Hochschule Hof zeigt auf, warum sich die deutschsprachige Wasserwirtschaft mit einer weiterentwickelten Digitalisierung immer noch sehr schwer tut. U.a. besteht auf Anwenderseite ein großer Wunsch nach Transparenz/Nachvollziehbarkeit, insbesondere wenn es um KI-basierte Lösungen geht (Müller-Czygan, Wimmer, Wagner, & Tarasyuk, 2021). Ein weiterer Aspekt ist die frühe Einbeziehung der Betriebsmitarbeiter, die in vielen Fällen entweder vernachlässigt oder sogar komplett unterlassen wird. Bereits im Jahr 2016 wurden im Forschungsprojekt KOMMUNAL 4.0 Lösungsan-

doi:10.18420/inf2022_125

sätze entwickelt, Erfahrungswissen zu erfassen und zu digitalisieren, um es in automatisierte Prozesse zu integrieren (Müller-Czygan, 2021). Dabei zeigte sich, dass gerade komplexe und multikriterielle Entscheidungsprozesse vom Betriebspersonal wesentlich detaillierter und auch in Bezug auf mögliche Vorhersageszenarien genauer beschrieben werden konnten. Das Erkennen von Muster mittels KI-Algorithmen verbessert sich zwar zunehmend, ist aber für Entscheidungen im Bereich der Umwelttechnik, wo falsche Entscheidungen verheerende Folgen haben könnten, noch zu ungenau. Insbesondere das Ableiten von Erkenntnissen und daraus erforderliche Handlungen erfordert in der Umwelttechnik weitaus mehr als ein schnelles berechnen möglicher Korrelationen. Hinzu kommt, dass gerade die kritischen und risikorelevanten Ereignisse komplex sind und nur vereinzelt auftreten und damit für ein selbstlernendes System zu wenig Lerndaten zur Verfügung stehen. Aus diesem Grund werden in der Wasserwirtschaft KI-Algorithmen für begrenzte Einsatzbereiche angewendet, die ein hohes Datenvolumen erzeugen (z.B. Daten aus einem biologischen Reaktorbecken mit mehreren Onlinemesssonden (Vijayaraghavan & Jayalakshmi, 2015) oder bei digitalen Wassermengenzählern in Privathaushalten (Rahim, Nguyen, Stewart, Giurco, & Blumenstein, 2020) bzw. bei frei verfügbaren Metadaten z.B. zur Prognose großflächiger Veränderungen der Wasservorkommen durch Trockenheit (Gino Sophia, Ceronmani Sharmila, Suchitra, Sudalai Muthu, & Pavithra, 2020)).

Auch wenn zunehmend der Zusammenhang zwischen den Betriebsbedingungen in einem Kanalnetz und der am Ende des Kanals liegenden Kläranlage digital erfasst werden kann, gibt es aktuell keine Stadt oder Gemeinde, die sowohl Kanal als auch Kläranlage vollständig mit Sensoren ausgestattet hat und damit über die notwendige Datendichte verfügt, die für ML-Analysemodelle erforderlich ist. In vielen Fällen werden Modellsimulationen mit ausgewählten Einzelmessungen gekoppelt (Mikkelsen, 2018), die dann die notwendige Datendichte erzeugen, um ML-Analysemodelle für Vorhersagen einzusetzen. Die Metastudie zum Stand der Digitalisierung in der deutschsprachigen Wasserwirtschaft der Hochschule Hof (Müller-Czygan, Wimmer, Wagner, & Tarasyuk, 2021) hat gezeigt, dass Betreiber komplexer wasserwirtschaftlicher Systeme selbst bei ausreichender Daten- und Algorithmentransparenz keine automatisierten Entscheidungen akzeptieren. KI-basierte Systeme sollen lediglich eine bessere Datenlage herbeiführen, auf dessen Basis die verantwortlichen Experten bessere, aber immer noch subjektive Entscheidungen treffen können, insbesondere in Krisensituationen wie z.B. bei Hochwasser oder Trockenperioden.

Das Cased Based Reasoning (kurz CBR) befasst sich mit einem sehr aktuellen Problem im Bereich der Künstlichen Intelligenz, und zwar mit der Entwicklung intelligenter (Experten-) Entscheidungsunterstützungssysteme, insbesondere in Echtzeit (real-time-ISEUS), die auf offene und dynamische Themenbereiche ausgerichtet sind. Es ist ein Ansatz, der es ermöglicht, ein neues, unbekanntes Problem zu lösen, indem man die Lösung eines bereits bekannten Problems verwendet oder anpasst, d.h. bereits gesammelte Erfahrungen bei der Lösung ähnlicher Situationen nutzt. Solche Systeme beruhen auf der In-

tegration von Modellen zur Anpassung, Änderung und zum Erlernen der Wissensdarstellung und der Funktionsweise, die sich an den Besonderheiten des Problem(sach)bereichs und der entsprechenden Art der Ungewissheit orientieren, was ihre Fähigkeit zur Entwicklung und Veränderung ihres Zustands widerspiegelt. Leider sind die meisten praktischen Probleme, die sich auf offene und dynamische Themenbereiche beziehen, schlecht formalisiert, und die Ungewissheit kann nicht-probabilistischer Natur sein. Bei der Suche nach einer Lösung für solche Probleme ist es notwendig, plausible Inferenzmethoden anzuwenden, die es ermöglichen, eine akzeptable (möglicherweise nicht optimale) Lösung zu finden. Ein Ansatz beruht auf der Tatsache, dass es für eine Person (Experte) charakteristisch ist, in der ersten Phase nach einer Lösung für ein neues (unbekanntes) Problem zu suchen, Lösungen zu verwenden, die bereits in ähnlichen Fällen gefunden wurden, und diese gegebenenfalls an das Problem (die aktuelle Problemsituation) anzupassen. Dieser erfahrungsbasierte Ansatz bildet die Grundlage für die Modellierung von Argumentationsmethoden auf der Grundlage von Präzedenzfällen. Für den Fall Heringhausen erscheint diese Form der Vorhersageanalyse das passende Mittel zu sein. Die wesentlichen Herausforderungen bestand darin, die relevanten Präzedenzfälle zu finden, technisch ausreichend zu beschreiben und das System für neue, noch nicht bekannte Fälle offen und anpassbar zu gestalten. Sobald die erforderlichen Fallbeschreibungen in der notwendigen Informationsqualität vorliegen, ist die Gestaltung mit dem CBR das vergleichsweise kleinere Problem.

Die Verwendung von Informationen bekannter hisotrischer und immer wiederkehrender Fälle mit CBR-Ansätzen werden in der Abwassertechnik seit rund zwei Jahrzehnten untersucht (z.B. (Krovvidy & Wee, 1993), (Wiese, Stahl, & Hansen, 2005)). Kombinationen mit Neuronalen Netzen und Fuzzy Logic wurden ebenfalls vorgenommen (Sarabia, et al., 2005). Die Anwendungen beschränkten sich aber oftmals auf Einzelfragen wie z.B. das Erkennen von ungewollten Mikroorganismen (Schmitt, Stahl, & Hansen, 2002), der Entfernung von Stickstoff (Quan, 2018), zur Vorhersage von Pumpenausfällen (Trstenjak, Palasek, & Trstenjak, 2019) oder als Hilfe beim Monitoring von Kläranlagen (Atanassov & Boshnakov, 2014). Die Kombination von CBR mit Simulationsmodellberechnungen und Echtzeitdaten aus einem SCADA-System als prognostische Entscheidungshilfe für eine Gesamtbetrachtung Kanal-Kläranlage mit dem Ziel eines energieeffizienten Gesamtbetriebs mit gleichzeitiger Optimierung des Reinigungsbetriebs vor dem Hintergrund stark veränderlicher Betriebsbedingungen erfolgt zumindest für Deutschland nach Kenntnis der Autoren erstmalig.

Der Weg zur Definition in Frage kommender Fälle für die Kläranlage Heringhausen, der auch hinsichtlich der Aussagequalität des CBR-Systems die größte Herausforderung dargestellt hat, wird im Folgenden beschrieben.

Klimawandel verändert Abwasserbedingungen – Lösung für Anpassung liegt in dynamischen Digitalisierungslösungen

Bisher werden fast alle Kläranlagen in Deutschland bei Mischwasserzufluss (Abwasser und Regenwasser werden in einem gemeinsamen Kanal zur Kläranlage abgeleitet) mit einem fest definierten Maximalzufluss zur Kläranlage betrieben. Überschreitet der Kanalabfluss diesen Maximalwert und ist das Speichervolumen im Kanalnetz erschöpft, kommt es zu sogenannten Entlastungsereignissen. D.h. verschmutztes Mischwasser wird meist ohne Behandlung an definierten Stellen direkt in das Gewässer abgeschlagen, obwohl oftmals die angeschlossene Kläranlage noch über genug Prozessreserve verfügen würde. Insbesondere in den Sommermonaten kann es bei kurzen, aber starken Regenfällen zu einer kurz- und mittelfristigen Anhebung chemischer Verschmutzungsparameter und einem Sauerstoffrückgang in Gewässern kommen, wenn ungereinigtes Mischwasser aus dem Kanal in ein Gewässer überfließt (Krieg, 2000). Unter Nutzung des Gesamtsystems Kanal-Kläranlage können Zuflussmengen mit Hilfe modernster Mess- und Automationstechnik dynamisch an die aktuelle stoffliche und hydraulische Belastungskapazität einer Kläranlage angepasst werden (Wiese, 2005). Insbesondere ist es dadurch möglich, auch die Anzahl der kritischen Entlastungsereignisse aus dem Kanalnetz, und damit unnötige Gewässerverschmutzungen zu reduzieren. Ein derartiges Konzept sollte daher auch auf der Kläranlage Heringhausen zum Einsatz kommen, da in den Sommermonaten aufgrund des touristischen Einflusses das Mischwasser tendenziell stärker verschmutzt ist. Aber auch bei Trockenwetter werden viele Kläranlagen in Deutschland nicht optimal betrieben, da sie in aller Regel mit einer fest eingestellten Steuerung ausgestattet sind, die nicht flexibel auf Belastungsschwankungen reagieren können. Auch in diesen Fällen kann durch eine dynamische, situations- und belastungsabhängige Steuerung eine deutliche Verbesserung erzielt werden. Zudem sollen moderne Systeme so eingesetzt werden, dass der Energieverbrauch auf ein erforderliches Minimum reduziert werden kann. Es sei darauf hingewiesen, dass Kläranlagen oftmals der höchste Einzelstromverbraucher einer Stadt oder Gemeinde sind mit bis zu 10 % am Gesamtstromverbrauch.



Abbildung 1: Kläranlage Heringhausen – aus optischen Gründen und für einen besseren Winterbetrieb wurde um die Anlage eine Halle gebaut (Bildquelle: J. Wiese)



Abbildung 2: Kläranlage Heringhausen – 1 = SB-Reaktor Nr. 1, 2 = SB-Reaktor Nr. 2, 3 = Überschussschlammspeicher (Bildquelle: J. Wiese)

Wie sieht die Realität aus? Identifikation und Beschreibung notwendiger Lastfälle

Im Zuge der Überlegungen für eine intelligente Steuerungslösung für die neue Kläranlage Heringhausen wurde mit dem Ziel der maximalen Algorithmentransparenz sowie der Einbeziehung der Betriebsmitarbeiter das CBR-Verfahren ausgewählt (Müller-Czygan, Wiese, & Vogel, 2021). Zudem hatte die Gemeinde Diemelsee als Betreiber der Kläranlage Heringhausen weder ausreichend historische Daten von Kanalnetz und Kläranlage, noch die finanziellen Mittel, im Zuge der Neubau- und Sanierungsarbeiten eine ausreichende Sensorenanzahl zu installieren, um die notwendige Datendichte zum Anlernen von KI-Modellen zu erzeugen. In Heringhausen liegt der spezielle Fall vor, dass die Abwasserzuflussbedingungen aufgrund des Tourismus extrem schwanken. Dabei ändern sich nicht nur die Abwassermengen je nach Saison, gerade die extreme Zunahme des Campingmobil-Tourismus stellt für attraktive Orte wie Heringhausen (mit dem Badesee Diemelsee) eine besondere Herausforderung dar. Campingmobile führen eigene Toiletten-/Abwassersysteme mit sich, die meist auf Stellplätzen wie am Diemelsee entleert werden. Neben einer hohen Konzentration an Abwasserschmutzstoffen enthalten die Sammelsysteme aus hygienischen Gründen zudem noch chlorhaltige Chemikalien, die für die biologischen Behandlungsprozesse eine starke Belastung bedeuten, wenn diese extremen Schmutzstöße der Kläranlage nicht vorausschauend zugeführt werden. In Heringhausen ist wie bei vielen Kläranlagen in Tourismusgebieten mit weitaus mehr unterschiedlichen Lastfällen zu rechnen als bei normalen Kläranlagen. Unabhängig des erforderlichen CBR-Algorithmus galt es, die in Frage kommenden Lastfälle zu ermitteln und mit Hilfe eines Simulationsmodelles die erforderliche Datendichte für einen passenden CBR-Algorithmus zu erzeugen. Realdaten in ausreichender Menge standen ja nicht zur Verfügung. So wurde zuerst ein detailliertes, computergestütztes Modell erstellt, welches dazu verwendet wurde, verschiedene Lastfälle zu simulieren (z. B. Saison, Nicht-Saison, Tagesgänge, Trockenperioden, Mischwasserzufluss). Auf den Ergebnissen aufbauend wurden (und werden noch) geeignete Betriebskonzepte entwickelt, die dann als Sollwert-/Einstellparameter in die Kläranlagensteuerung eingehen, um adäquat auf die jeweilige Belastung reagieren zu können.

Das Modell bot nun die Möglichkeit, Eingangsdaten für die geplante CBR-basierte Entscheidungsunterstützung zu generieren. Zum Einsatz kommt das Simulationswerkzeug SIMBA#, das in Deutschland in der Abwasserwirtschaft weit verbreitet ist (Alex & Ogurek, 2020). SIMBA# enthält eine hohe Anzahl verschiedener Reinigungsmodelle und virtueller technischer Komponenten, sodass die verfahrenstechnische Seite einer Kläranlage vollständig beschrieben werden kann. Mit dieser Software können neben den biologischen Prozessen auch Energie- und Stoffströme simuliert werden, so dass auch das Ziel einer maximalen Energieeffizienz betrachtet werden kann. Die ungewollten Schmutzfrachtausträge aus Kanalnetz (inkl. Regenüberläufen) und Kläranlage können in SIMBA# gemeinsam simuliert werden. Das Modell liefert für weitere Untersuchungen auch „virtuelle (Online-) Messdaten“ für Entwicklung und Test geplanter Steuerungsalgorithmen.

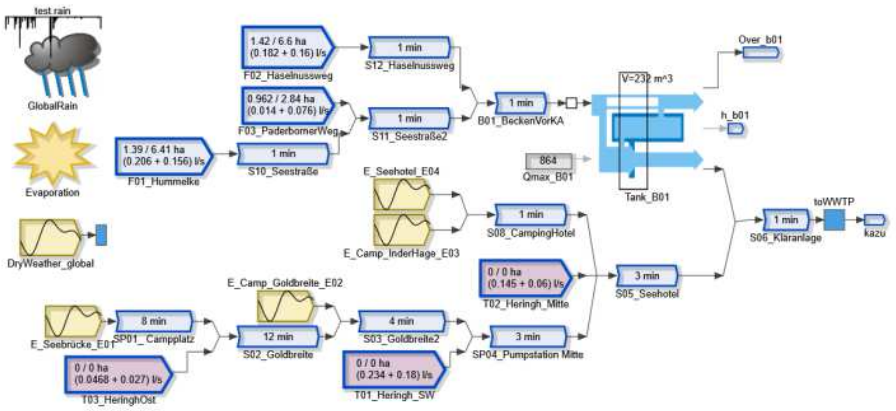


Abbildung 3: SIMBA#-Kanalnetzmodell des Ortsteils Heringhausen (Quelle: eigene Darstellung)

Abbildung 3 zeigt beispielhaft das mit SIMBA# erstellte Kanalnetzmodell der Einzugsgebiete der Kläranlage Heringhausen für einen Belastungsfall aus dem Simulationsprogramm SIMBA#. Die einzelnen Symbole stellen die für den Lastfall relevanten Abflussmengen dar, die als Regen- und Abwassergemisch (blau), als reines Abwasser (gelb) und als sogenanntes Fremdwasser (violett, z.B. Grundwasser, welches in undichte Kanäle eintritt) der Kläranlage zufließen. Die dahinerliegenden Berechnungsschritte erlauben es, die Auswirkung unterschiedlicher Wetter- und Zuflussereignisse auf den Zulauf zur Kläranlage mit eingegebenen Niederschlagsprognosen abschätzen. Prognosen sind abhängig von Intensität und Dauer des Niederschlags, von der Verteilung auf die Einzugsgebiete und den Fließzeiten im Kanalnetz und den Kapazitäten vorgeschalteter Stauräume. Mit dem Computermodell SIMBA# ist dies alles erfassbar, so lassen sich somit viele verschiedene Belastungssituationen für eine Kläranlage testen (Eingabe unterschiedlicher relevanter Abflussmengen je nach Wetterdateneingabe und erwarteter Touristenanzahl) und als Basis für ein CBR-System erzeugen.

Ausgehend von den Zuflussszenarien des Kanalnetzmodells (Abbildung 3) erfolgt im nächsten Schritt die Untersuchung, wie sich die verschiedenen Zuflussszenarien auf die Performance der Kläranlage auswirken können. Für die Aufgabenstellung der Kläranlage Heringhausen wurde mit dem Ziel eines adäquaten Verhältnisses von Aufwand und Nutzen ein entsprechendes Modell der Kläranlage durch grafische Verschaltung von Funktionsblöcken entsprechend Abbildung 4 erstellt. Abgebildet ist das in Heringhausen angewandte sogenannte SBR-Verfahren (SBR = Sequenced Batched Reactor), eine Verfahrensart der biologischen Abwasserreinigung, die besonders für den energieeffizienten Einsatz mit schwankenden Zuflussmengen geeignet ist (Wiese, 2004).

Das aus dem Kanal kommende Abwasser fließt nach erfolgter Abtrennung von Grobschmutzstoffen zuerst in einen Vorlagebehälter, der dazu dient, die Zulaufspitzen zu dämpfen und für eine gezielte Abwasserzugabe in die Reaktoren (SBR 1 und 2) zum Abbau organischer Schmutzstoffe zu sorgen (biologische Behandlung). Das Modell enthält wie die echte Anlage alle realisierten Pumpen, Rührwerke und Messsensoren der jeweiligen Prozessstufen (als Funktionsblöcke ausgewiesen) sowie die in Realität programmierten Steuerungsprozessschritte und bildet somit den Abwasserprozess nahezu in realgetreuer Weise ab.

Abbildung 5 zeigt exemplarisch den Baustein für die Steuerung der beiden biologischen-SBR-Reaktoren. Die in der biologischen Stufe stattfindenden mikrobiologischen Prozesse werden mathematisch beschrieben, deren Inputdaten dargestellt und die zu erwartenden Outputdaten ermittelt. Maßgebend für die Bewertung der gesamten Simulationsberechnungen sind die Outputdaten, die mit den in der Realität einzuhaltenden Schmutzstoffparameter übereinstimmen, die eine Genehmigungsbehörde dem Kläranlagenbetreiber als Einleitgrenzwerte in ein Gewässer vorschreibt. Wichtige verfahrenstechnische Prozesse wie die Zugabe von Prozessluft, die Abwassermengenzugabe oder die Durchmischung des Abwassers mittels Rührwerke werden mathematisch ebenso berücksichtigt wie das Wachstum der Abwasserbakterien sowie die Entnahme der überschüssigen Bakterienmasse. Von besonderer Bedeutung sind die aus dem Simulationsmodell zu generierenden Steuerungsparameter. Sobald ein Lastfall die gewünschten Outputdaten erzeugt (z.B. Konzentration organischer Substanzen, Stickstoff- und Phosphatgehalt oder pH-Wert im gereinigten Abwasser), wird ermittelt, mit welchen Steuerungsparametern dies erreicht wurde (z.B. Lufteintragsmenge, Vermischungsenergie, Abwasserzugabemenge). Sie sind die entscheidenden Daten, die der Abwasserbetrieb in Heringhausen benötigt, um die reale Anlage so einzustellen, dass auch der reale Prozess gemäß den behördlichen Auflagen betrieben wird.

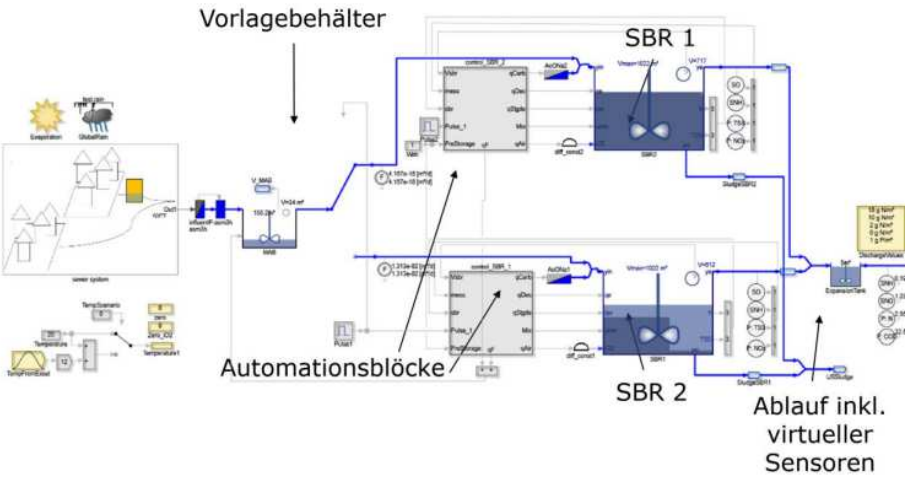


Abbildung 4: SIMBA#-Kläranlagenmodell des Ortsteils Heringhausen (Quelle: eigene Darstellung)

Die biologischen Prozesse wurden nachgebildet auf Basis des sog. Activated Sludge Modell No. 3 (Koch, Kühni, Gujer, & Siegrist, 2000), einem seit vielen Jahren weltweit bewährten mathematischen Modell zur Nachbildung des Kohlen- und komplexen Stickstoffstoffabbaus biologischer Abwasserprozesse, welches in Abbildung 6 als Anschauung dargestellt ist. Zur weiteren Erklärung wird auf die einschlägige Literatur verwiesen.

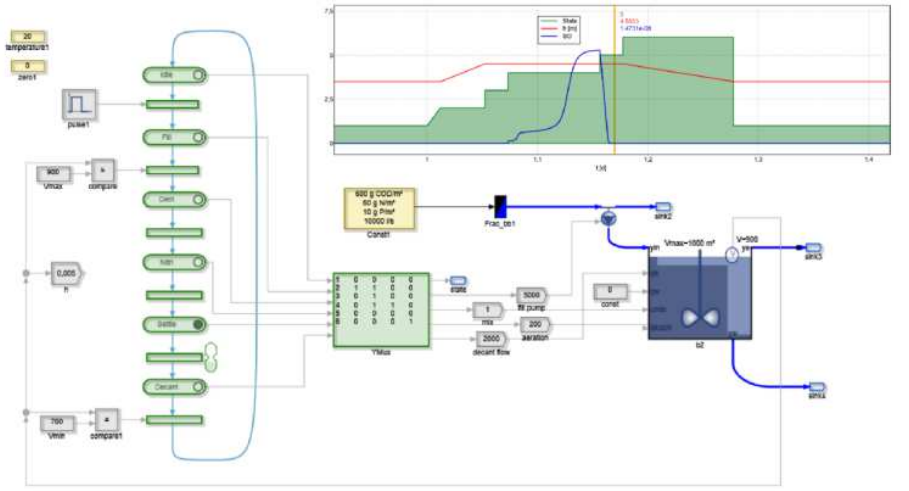


Abbildung 5: SIMBA#-Modell der Zyklensteuerung (Quelle: eigene Darstellung)

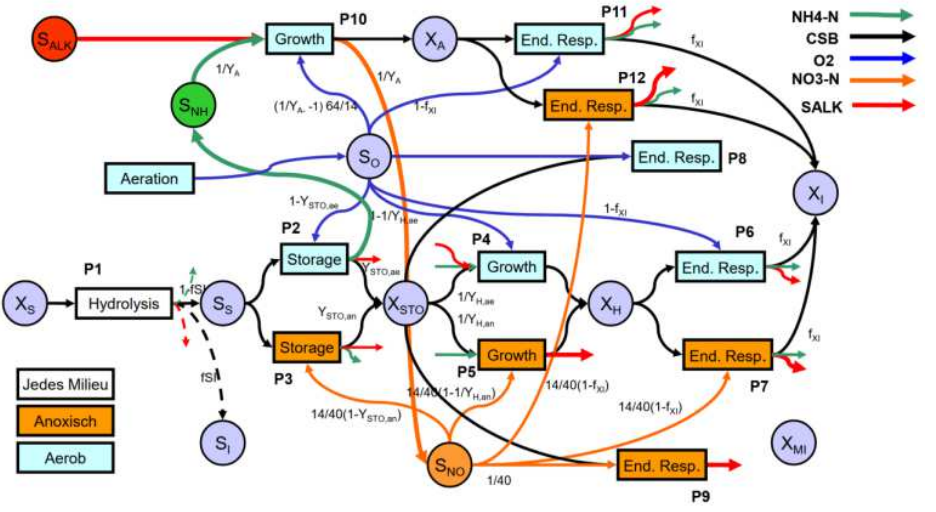


Abbildung 6: Prinzip des verwendeten biologischen Modells zum Schmutzstoffabbau (IFAK, 2022)

Die CBR-Lösung der Kläranlage Heringhausen vereint Transparenz, Betriebswissen und überlässt die verantwortliche Entscheidung dem Menschen

Unter Einbezug des Betriebspersonals wurden zahlreiche verschiedene Belastungssituationen (Trocken-/Regenwetter, unterschiedliche Abwassertemperaturen, Normalbelastung/ Ausbaubelastung/Extrembelastung etc.) definiert, die mit Hilfe des zuvor beschriebenen SIMBA#-Kläranlagenmodells berechnet und überprüft wurden. Hierbei spielte das Erfahrungswissen des Betriebspersonals die entscheidende Rolle, da keine ausreichende Anzahl dokumentierter Daten vorlag, was typisch ist für Kläranlagen im ländlichen Raum. Über das gesamte Betriebs- und Prozesswissen verfügt fast ausschließlich das Betriebspersonal und ggf. beauftragte Ingenieurbüros, die eine solche Anlage geplant und betreut haben. Je stärker und detaillierter das Betriebspersonal sein Erfahrungswissen in die Modellerstellung als Basis der CBR-Berechnung einbringen durfte, desto engagierter wurde es insgesamt bei der Implementierung innovativer digitaler Lösungen. Eine mangelnde Akzeptanz der späteren CBR-Lösung wurde nicht mehr festgestellt. Diese Erfahrungen decken sich mit den Ergebnissen zur Akzeptanz digitaler Lösungen durch Einbezug von Betriebswissen aus der Metastudie der Hochschule Hof zum Stand der Digitalisierung der deutschsprachigen Wasserwirtschaft (Müller-Czygan, Wimmer, Wagner, & Tarasyuk, 2021).

Die intensive Zusammenarbeit zwischen Modellentwickler und Betriebspersonal hat gezeigt, dass besonders in Heringhausen eine hohe Variabilität des gesamten Abwasserreinigungsprozesses notwendig ist, da eine hohe Anzahl unterschiedlicher Zulaufbedingungen (Lastfälle) zu berücksichtigen waren. Insgesamt wurden 20 verschiedene relevante Lastfälle identifiziert, mit deren Auftreten heute und in Zukunft zu rechnen ist. Für alle 20 Fälle wurden mit dem Betriebspersonal die Inputdaten ermittelt und mit dem beschriebenen Simulationsmodell die Outputdaten. Als wesentliches Ziel der Lastfallsimulationen galt es, für jeden einzelnen Lastfall die erforderlichen Steuerungsparameter zu ermitteln, damit das Betriebspersonal bei Auftreten eines Lastfalls die Kläranlage optimal einstellen kann. Tabelle 1 zeigt die zum Zeitpunkt der Publikationserstellung bereits durchgeführten und ausgewerteten 8 Lastfallsimulationen. Die weiteren 12 identifizierten und definierten Lastfälle werden derzeit analysiert. Anschließend werden die zugehörigen optimalen Steuerungsparameter generiert.

Aufgrund des fortschreitenden Klimawandels ist damit zu rechnen, dass neue, bislang nicht identifizierte bzw. definierte Lastfälle auftreten. Auch eine Veränderung der Tourismusentwicklung am Diemelsee kann zu neuen Lastfällen führen. Um in Zukunft diese weiteren Lastfälle betrachten zu können, wurde unter SIMBA# ein spezieller Zulaufgenerator entwickelt, mit dem verschiedene Zulaufkombinationen und Zulaufsituationen erfasst und für die weiteren Simulationen gemäß Abbildung 4 und 5 erstellt werden können.

Tabelle 1: Bereits durchgeführte und ausgewertete 8 Lastfallsimulationen (Quelle: eigene Darstellung)

Nr.	Zulaufteilströme: Schmutzwasser+ Fremdwasser+ Regen	Mengen [m³/d]	Temp. °C	Regen	Kommentar, Anmerkung zum Zulauf	Bewertung, z.B. simulierte Ablaufwerte [g/m³]
1	480 EW + 1500 EW Gewerbe Fremdwasser	72 +225 +700 = 997	20	Ohne	Basis-TW im Sommer	Keine Probleme, NH_4-N < 0,3 NO_3-N < 2,2
2	480 EW + Fremdwasser	72 700 = 772	10	Ohne	TW Minimum Nur EW+ viel FW	Schwachlast
3	480 EW + 1500 EW Gewerbe Fremdwasser	72 225 700 = 997	10	Ohne	Basis TW im Winter	NH_4-N < 0,3 NO_3-N < 4
4	Basis TW + Fremdwasser+ Regen	297 700 700 =1.697	10	Blockregen	BasisTW mit viel Regen verdünnt	Hydraul. Überlastung des MAB, Anpassung des Betriebsregimes, verkürzter Zyklus!
5	TW 1.980 EW (Q_{max} 1.296 m³/d) + max. Zufluss Mischwasser 1.996 m³/d (davon Regen)	997 300	10	Regenreihe, bzw. Blockregen	2. Auslegungs-Lastfall,	Anpassung des Betriebsregimes erforderlich
6	TW 3.000 EW Schmutzwasser + Fremdwasser (83,17m³/d)	450 700	10	Ohne	Maximallast	NH_4-N < 1,5 NO_3-N < 7
7	TW 3.000 EGW Schmutzwasser + Fremdwasser +Regen	450 250 300	10		Maximallast mit Regen verdünnt	Hydraulische Überlastung des MAB
8	TW 1.980 EW + 3500 EW zusätzlich Wenig FW	297 525 0	20		Zusätzliche 4000 Übernachtungsgäste +1000 Tagesgäste	Im Sommer NH_4-N < 1,5 NO_3-N < 7,5

Die alleinige Beschreibung und Berechnung von Outputdaten und Steuerungsparametern möglicher Lastfälle mit einem Simulationsprogramm wie SIMBA# hat für das Betriebspersonal von Abwasserinfrastruktureinrichtungen keinen großen Handlungswert. Zwar liegen nun Informationen vor, wie Kanal und Kläranlage für jeden Lastfall energieeffizient und rechtssicher zu betreiben sind. Da sich die Auswirkungen von Steuerungsanpassungen in biologischen Abwassersystemen erst nach vielen Stunden bis Tagen zeigen, ist ein rechtzeitiges Handeln und ein vorausschauender Systembetrieb wünschenswert. Die optimale Nutzung von Simulationsinformationen ist dann gegeben, wenn ein permanenter Abgleich der Lastfalldaten mit der Realsituation erfolgt, und ein weiteres System diese Daten auswertet und dem Betrieb aufzeigt, mit welchen Lastfällen in naher Zukunft zu rechnen

ist, um komplexe Abwassersysteme rechtzeitig auf neue Lastfälle anzupassen. Bisherige Versuche mit neuronalen Netzen oder Fuzzy-Logiken wurden in der Abwassertechnik zwar erprobt und für Einzelprozesse wie z.B. die Luftregelung auch realisiert. Deren Anwendung ist bei komplexen Gesamtsystemen aber nicht gewünscht, da diese Lösungen keine ausreichende Transparenz der Ergebnisermittlung bieten (Müller-Czygan, Wimmer, Wagner, & Tarasyuk, 2021). Aus diesem Grund wurde für das Vorhaben Heringhausen der Einsatz des CBR-Verfahrens untersucht.

Wie bereits erläutert, handelt es sich bei dem CBR-Verfahren um einen Ansatz, der es ermöglicht, ein neues, unbekanntes Problem zu lösen, indem man die Lösung eines bereits bekannten Problems verwendet oder anpasst.

Mit Hilfe der beschriebenen Kanal- und Kläranlagenmodellbildung in SIMBA# wurden die benötigten Modelle zur Anpassung, Änderung und zum Erlernen der Wissensdarstellung und der Funktionsweise der Kläranlage Heringhausen inkl. zugehörigem Kanalnetz erstellt. Das Ziel ist es nun, mit Hilfe eines eigens entwickelten CBR-Algorithmus Realität und Erfahrungswissen (in Form der modellierten Lastfälle) zu vergleichen, so dass die Anlagenbedienenden in den unterschiedlichen Wetter- und Zuflusssituationen über das CBR-System einen Vorschlag bekommen, welcher Modelllastfall oder welche Modelllastfälle der aktuellen Realität an ehesten entsprechen und welche Steuerungsparameter sich daraus für den Kläranlagenbetrieb ergeben. Dazu werden die aktuellen Echtzeitdaten aus dem vorhandenen SCADA-System (Monitoringsoftware für das gesamte Abwassersystem) sowie die mit SIMBA# ermittelten Lastfalldaten zur CBR-Berechnung herangezogen (siehe Abbildung 7). In der CBR-Berechnung werden die Auswahlparameter der Lastfälle mit den Realwerten verglichen und in entsprechend festgelegte mathematische Beziehungen gesetzt. Hierbei werden insbesondere die Zulaufdaten (Menge und Schmutzfracht), die Wetterdaten und die aktuelle Touristenauslastung analysiert und mit den charakteristischen Parametern der Lastfälle verglichen. Werden die definierten Übereinstimmungen festgestellt, gilt ein identifizierter Lastfall als gültig und wird dem Betriebspersonal zur Anwendung vorgeschlagen. Da gerade in der Anfangsphase eine gewisse Unsicherheit beim Betriebspersonal herrschen wird, sollen ergänzend auch Lastfälle mit etwas geringerer Übereinstimmung als definiert angezeigt werden, so dass das Betriebspersonal eine größere Fallauswahl hat und daran lernt, wie es dem System vertrauen kann. Im Gegensatz zu dynamischen CBR-Systemen erzeugt das Heringhauser-CBR-Modell keine eigenen neuen Lastfallkombinationen, sondern erstellt nur den Soll-Ist-Vergleich zwischen den Lastfallmöglichkeiten und der Realität.

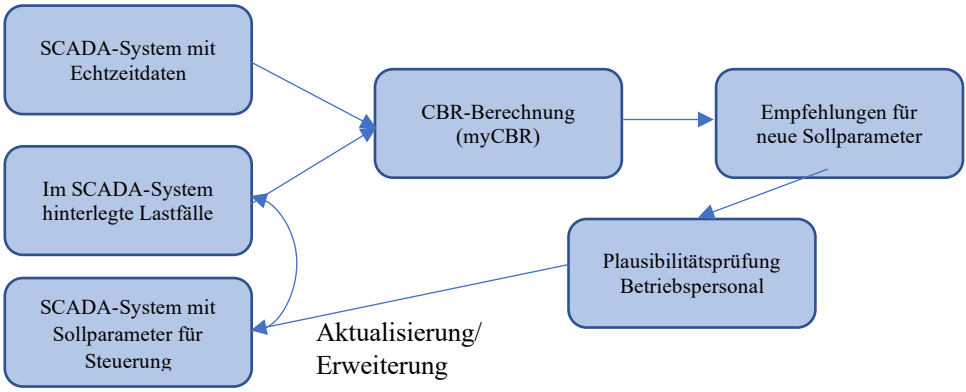


Abbildung 7: Soll-Ist-Vergleich mit CBR-Berechnung

Da das Heringhauser-CBR-Modell exemplarisch ist für die meisten Kläranlagen in Deutschland (von den ca. 10.000 Kläranlagen in Deutschland sind ca. ein Drittel in einer vergleichbaren Größenordnung wie Heringhausen), und das dort tätige Personal vorwiegend gewerblich ausgebildet ist, ist eine Funktionsbeschränkung auf einen Soll-Ist-Vergleich zwischen Lastfallmöglichkeiten und Realität auch aus Akzeptanzgründen zielführend.

Damit wird auch eine Schwäche der Ursprungsversion des CBR-Vertfahrens begrenzt, die in der begrenzten Fähigkeit liegt, Regeln zu schaffen bzw. die Erfahrungen zu verallgemeinern oder eigene Klassifizierungen zu schaffen. Durch die Festlegung definierter Lastfälle (und damit auch Klassifizierungen) ist diese Funktion für das Heringhauser-CBR-Modell auch nicht notwendig und erfordert zudem die umfassende Beteiligung des Betriebspersonals, was ein hohes Akzeptanzmotiv und ein wesentlicher Erfolgsfaktor für eine erfolgreiche Umsetzung darstellt. Abbildung 8 stellt den vorgesehenen CBR-Zyklus dar, der softwaretechnisch umzusetzen ist.

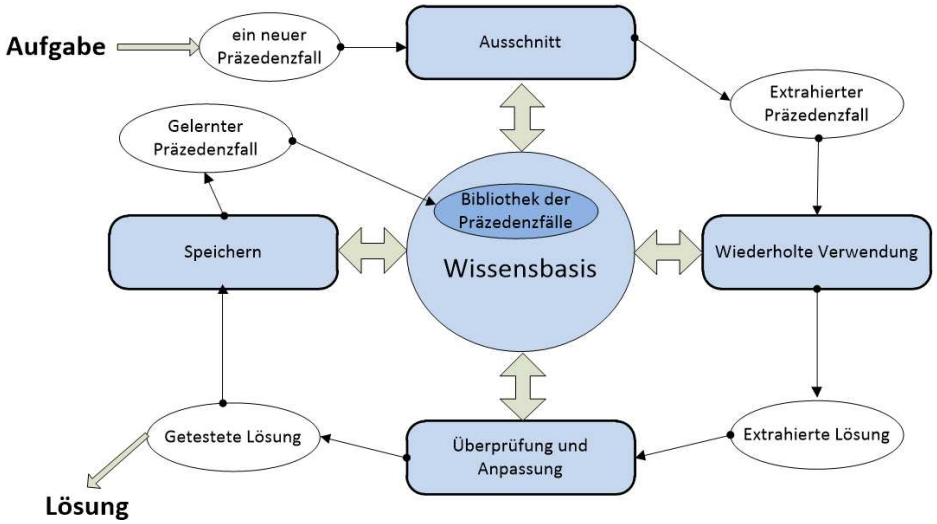


Abbildung 8: Prinzip CBR-Zyklus

Die Auswahl des CBR-Verfahrens für eine KI-basierte Entscheidungshilfe in der Abwassertechnik am Beispiel von Kanalnetz und Kläranlage in Heringhausen gegenüber Verfahren, die eine hohe Datendichte und umfangreiche Datenhistorie benötigen, liegt in den systembedingten Vorteilen des CBR:

- die Möglichkeit, die auf Basis einmal gesammelten Erfahrungen des Betriebspersonals, die in entsprechenden Lastfalldefinitionen überführt sind, direkt zu nutzen, ohne dass für neue oder unklare Situationen ein Experte für ein bestimmtes Fachgebiet hinzugezogen werden muss;
- die Möglichkeit, die Lösungszeit zu verkürzen, indem eine bereits vorhandene Lösung für ein ähnliches Problem verwendet wird;
- die Möglichkeit, die wiederholte Erlangung einer fehlerhaften Lösung auszuschließen;
- keine Notwendigkeit für ein gründliches Studium und die Nutzung des gesamten verfügbaren Wissens über das Fachgebiet, da nur die wesentlichen Merkmale (in diesem Fall relevanten Parameter und Zustandsbeschreibungen) des Fachgebiets berücksichtigt werden können;
- die Möglichkeit, Heuristiken zu verwenden, die die Effizienz des Lösungsfindungsprozesses erhöhen.

Das CBR-Verfahren ist generell auch für laienhafte (grobe) Lastfallbeschreibungen geeignet. Wird es in der Abwasserwirtschaft verwendet, muss aufgrund der komplexen Anforderungen darauf geachtet werden, dass die Lastfallbeschreibungen so exakt und so nah an

der Realität wie möglich sein, um vor dem Hintergrund möglicher Umweltauswirkungen negative Folgen zu vermeiden.

Vom Konzept zum CBR-Technologiedemonstrator

Durch erhebliche Verzögerungen beim Bau und Inbetriebnahme der Kläranlage Heringhausen (infolge der Covid-Pandemie) lagen zu Beginn der CBR-Entwicklung nur wenige belastbare Betriebsdaten der Neuanlage vor. Aus diesem Grund wurde mit dem Auftraggeber abgestimmt, dass ein erster CBR-Technologiedemonstrator auf die die zuvor beschriebenen Lastfälle der SIMBA#-Simulation zurückgreift. Im Laufe der nächsten Jahre können durch den Auftraggeber weitere reale Lastfälle in das System eingepflegt werden, sodass sich dann die Datenbasis und die Qualität der Prognose sukzessive verbessert. Es gibt zudem bereits Überlegungen, das CBR-Modell direkt in die SCADA-Umgebung zu integrieren.

Aktuell liegt der CBR-Technologiedemonstrator als Offline-Version vor. Die softwaretechnische Umsetzung erfolgte auf Basis der Software „myCBR“ in der Softwareumgebung „Protégé“ (kostenlos verfügbare Open-Source-Software (DFKI, 2022)).

Die Modellerstellung erfolgte nach den folgenden Schritten gem. (Chabroswki, 2022):

1. Implementierung von myCBR in Protégé
2. Informationsgewinnung über relevante Parameter
3. Einspeisung der Daten in myCBR
4. Kalibrierung der Fälle

Um das System am Anfang möglichst einfach zu halten, wurde die Ähnlichkeitssuche auf die folgenden Parameter beschränkt, die vom Betreiber ohne großen Aufwand eingegeben werden können:

- Aktuelle Zulaufmenge [m³ /d] – Diese Information stellt das SCADA-System in Echtzeit zur Verfügung
- Ferientag/Feiertag [-] – Diese Information ist dem Betriebspersonal (z.B. aus dem Kalender) ebenfalls bekannt
- Abwassertemperatur [˚C] - Diese Information stellt das SCADA-System in Echtzeit zur Verfügung
- Wochentag [-] – Diese Information ist dem Betriebspersonal (u.a. aus dem Kalender) ebenfalls bekannt

Die aktuelle Zulaufmenge lässt indirekt Rückschlüsse auf die aktuelle Schmutzstoffbelastung zu (d.h. Regenwetter, Fremdwasseranfall, Tourismus), da es hierzu lineare Zusam-

menhänge gibt. Da der Diemelsee ein Naherholungsgebiet ist, steigt die Belastung an Feiertagen und Ferientagen erfahrungsgemäß stark an; hingegen ist die Belastung an normalen Wochentagen geringer. Gleichwohl ist die Belastung nicht an allen Werktagen gleich, da der Trend zum verlängerten Wochenende dazu führt, dass die Belastungen an Montagen und Freitagen erfahrungsgemäß höher ist als an einem Dienstag, einem Mittwoch oder einem Donnerstag. Die Abwassertemperatur hat einen großen Einfluss auf die biologischen Prozesse und damit auf die Reinigungsleistung. Eine Übersicht über die in myCBR integrierten Fälle zeigt Abbildung 9.

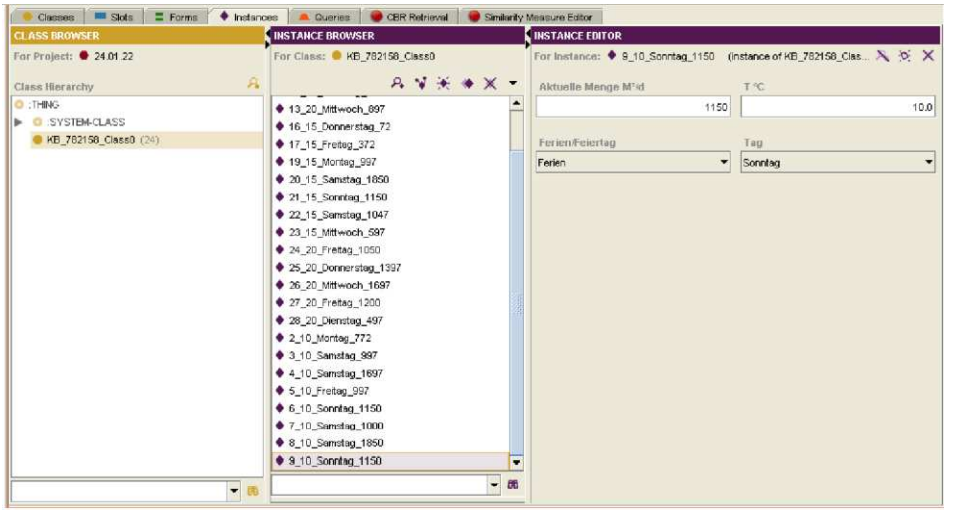


Abbildung 9: Fallübersicht „KA Heringshausen“ in myCBR (Bildquelle: (Chabroswki, 2022))

Im nächsten Schritt wurde damit begonnen, Ähnlichkeitsmerkmale zu definieren, damit geeignete Fälle aus der Datenbank identifiziert werden können. Hierzu kamen verschiedene Ansätze zur Erkennung von Ähnlichkeiten (Similarity Measure Functions) zur Anwendung. Wurde bei den Wochentagen mit einer Matrixähnlichkeit (Abbildung 10) gearbeitet, so kam der Abwassertemperatur und der Zulaufmenge ein Ansatz nach dem Euklidischen Distanzmaß (Abbildung 11) zur Anwendung („Ein Wert der nahe am Zielwert ist, wird höher gewichtet, als ein Wert der weiter entfernt vom Zielwert liegt“). Für das Ähnlichkeitskriterium „Ferien-/Feiertag“ reichte ein einfacher Abgleich (Ja/Nein). Nach diesen Vorarbeiten kann mit der Identifizierung ähnlicher Fälle begonnen werden (Case Retrievel). Ein beispielhaftes Case Retrievel ist in Abbildung 12 dargestellt: Der Abfragefall (Query), ein Samstag außerhalb der Ferienzeit mit einem Tageszufluss von 450 m³/d und einer Abwassertemperatur von 20 °C, führt dazu, dass aus der Datenbank der Fall „28_20_Dienstag_497“ als ähnlichste Lösung herausgesucht wurde. Der Zufluss lag im

Vergleichsfall mit 497 m³/d nur ca. 10 % neben dem Vorgabewert, bei den Ähnlichkeitswerten „Ferientag/Feiertag“ bzw. Abwassertemperatur gab es eine Übereinstimmung, lediglich beim Wochentag gab es Unterschiede (Samstag vs. Dienstag); die beiden anderen Fälle waren hingegen von der Ähnlichkeit deutlich schlechter.

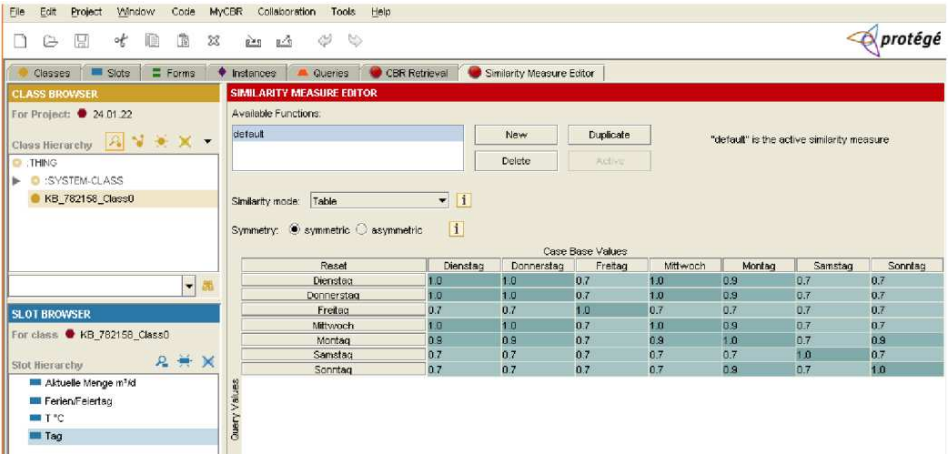


Abbildung 10: Similarity Measure Function für das Ähnlichkeitskriterium „Wochentag“ in myCBR (Bildquelle: (Chabroswki, 2022))

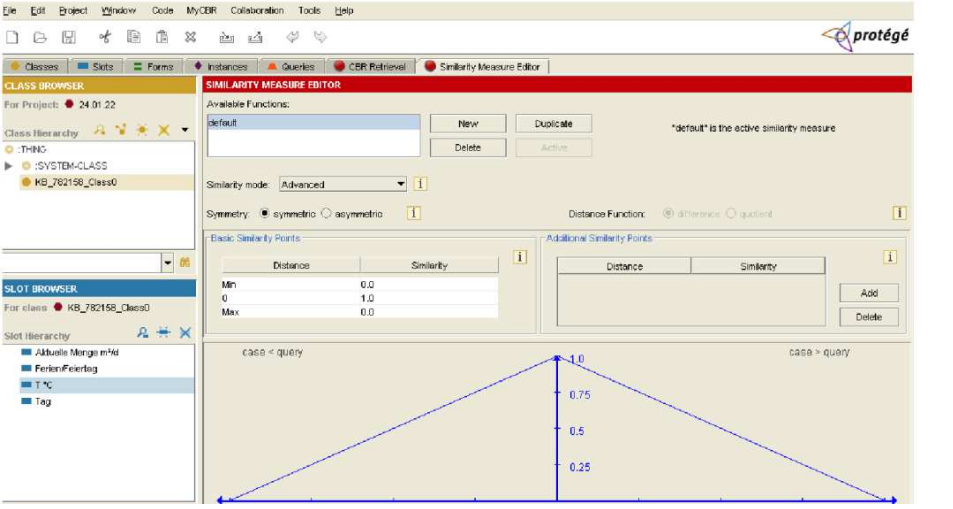


Abbildung 11: Similarity Measure Function für das Ähnlichkeitskriterium „Abwassertemperatur“ in myCBR (Bildquelle: (Chabroswki, 2022))

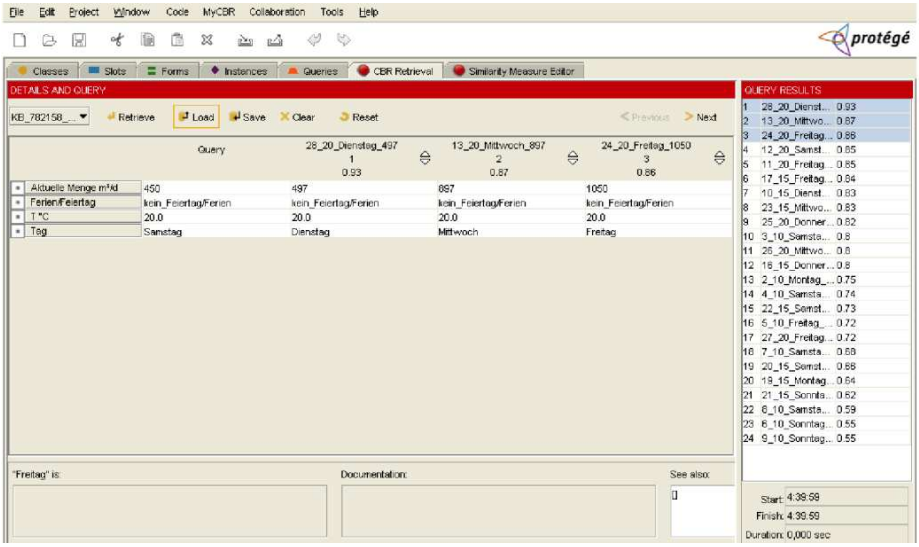


Abbildung 12: Exemplarisches Case Retrievel (Bildquelle: (Chabroswki, 2022))

Zug um Zug werden im weiteren Projektverlauf die derzeit in Analyse befindlichen Lastfälle in myCBR integriert und der CBR-Algorithmus optimiert.

KI muss zieldienlich sein und den Menschen mitnehmen

Auf Basis von myCBR wurde ein einfaches CBR-Modell erstellt, welches es in der ersten Projektphase anhand vier einfach zu ermittelnder Kriterien erlaubt, neue Fälle mit ähnlichen Fällen aus der Datenbank zu vergleichen. So wurden mit dem Betriebspersonal verschiedene Lastfälle identifiziert und definiert (siehe Tabelle 1). Da diese Fallbasis für alle relevanten Realitätsvarianten noch nicht ausreichend ist, wird in einer weiteren Projektphase die Fallbasis für das CBR-Modell noch einmal erhöht (z. B. durch Variation der Abwassertemperaturen). Perspektivisch kann in Zukunft die Fallbasis durch reale Betriebsdaten erweitert bzw. schrittweise ersetzt werden. Für die aktuell vorliegende Fallbasis wurde das SCADA-System bereits mit einer Falldaten- und Sollwertparametermatrix gemäß Abbildung 13 erweitert. In dieser Matrix sind alle für die Steuerung erforderlichen Sollparameter, die als Ergebnis der bisher vorliegenden Lastfallsimulationen in SIMBA# errechnet wurden, bereits eingetragen.

Die in der Spalte „aktuelle Sollwerte“ aufgeführten Parameter stellen die Eingriffsmöglichkeiten in die Steuerung der Kläranlage dar. Dies betrifft für den SBR-Abwasserprozess typischen Parameter wie z.B. die Abwasserzuführungsmenge, Länge von Zykluszeiten und Dauer/Intensität Luftzugabe, Dauer der Absetzzeit von Abwasserschlamm, Entnahmemenge überschüssiger Bakterienmasse oder die Grundeinstellung, ob sich der Reinigungsprozess an aktuellen Abwasserparametern automatisch ausrichtet oder nach starren Vorgaben abläuft.

Wie beschrieben, werden die zu berücksichtigenden Sollwerte (Steuerungsparameter) durch die Simulationsergebnisse der Lastfallanalysen vorgegeben. Das CBR-Modell ermittelt den der Realität am nächsten liegenden Lastfall für den betrachteten Zukunftszeitraum (Tage oder Wochen). Wiederkehrende Lastfälle, die an feste Termine gebunden sind wie Ostern, Sommerurlaub oder spezielle Feiertage mit typischem Touristenaufkommen können fix hinterlegt und über den Kalender abgerufen werden. Insbesondere die vom Wetter abhängigen Lastfälle sind permanent mit der Realität abzugleichen.

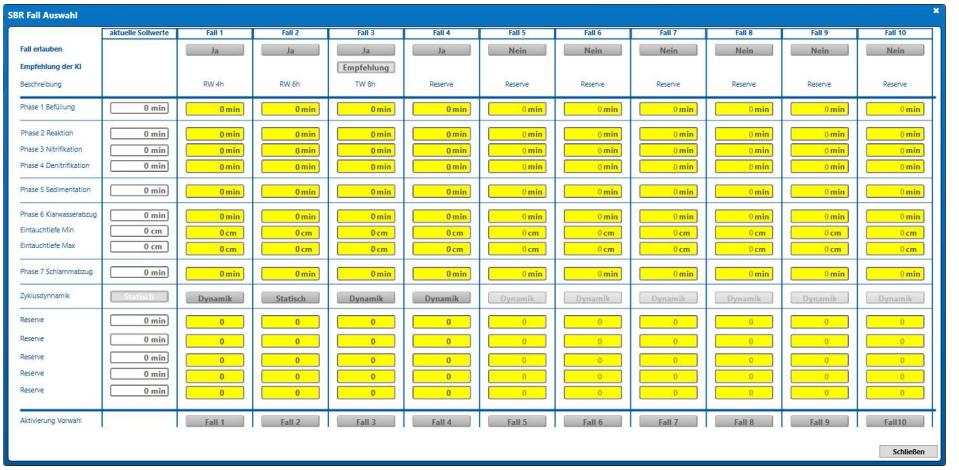


Abbildung 13: CBR-Falldaten- und Sollwertparametermatrix im SCADA-System der SBR-Kläranlage Heringhausen (Quelle: HST Systemtechnik GmbH & Co. KG)

Sukzessive werden die noch fehlenden Daten der 12 Lastfälle ergänzt. Wie bereits angesprochen, erhält das Betriebspersonal in der ersten Umsetzungsphase aus myCBR das Ergebnis der Lastfallanalyse als Entscheidungshilfe. Nun kann das Betriebspersonal den vorgeschlagenen Lastfall prüfen und bei Freigabe die Schaltfläche in der SCADA-Oberfläche (siehe Abbildung 13) auf „Ja“ setzen. In einem nächsten Schritt sollen die Ergebnisse aus

myCBR automatisch in das SCADA übertragen werden, und auf der SCADA-Oberfläche erscheint der Schriftzug „Empfehlung“, so dass myCBR nur im Hintergrund abläuft.

Wie wird es weiter gehen in Heringhausen und anderen Anwendungsfällen?

Nachdem ein Computermodell von Kanalnetz und Kläranlage des Ortsteils Diemelsee-Heringhausen erstellt wurde, wurden zahlreiche Simulationsläufe zu den verschiedensten Belastungsszenarien durchgeführt. Die Simulationsergebnisse deuten darauf hin, dass die Kläranlage Heringhausen unter Anwendung der Fallanalyse nach dem CBR-Verfahren in allen zu erwartenden Lastfällen sehr gute Reinigungsleistungen mit einem energieeffizienten Betrieb erzielen wird und damit erheblich zum Schutz des Badegewässers Diemelsee beiträgt. Es zeigte sich aber auch, dass unter widrigen Belastungsbedingungen (z. B. hohe hydraulische und stoffliche Belastung) die reale Steuerung noch Optimierungspotential besitzt. Dieses Optimierungspotential soll in naher Zukunft genauer identifziert und genutzt werden.

Auf Basis der Open Source-Software myCBR wurde bislang ein erstes offline-CBR-Modell erstellt, das anhand vier einfach zu ermittelnder Kriterien neue Fälle mit ähnlichen bekannten Fällen vergleichen kann. Im weiteren Projektverlauf gilt es nun die Erkenntnisse aus der Computermodellierung und insbesondere der Optimierung der Steuerung auf das CBR-Modell zu übertragen, sodass das System den Betreiber bei der Auswahl einer für einen spezifischen Fall geeigneten Steuerungsparametrierung unterstützen kann.

Im Zuge der Verbesserung der Abwasserinfrastruktur am Standort Heringhausen werden in 2023 auch die sogenannten Außenbauwerke wie Pumpstationen und Speicherbecken weiter digitalisiert. Hier werden zusätzliche Daten generiert, die in der Kläranlagensteuerung berücksichtigt werden sollen bzw. zur Generierung weiterer Lastfälle heranzuziehen sind. Derzeit arbeiten diese Bauwerke autark, d.h. ob beispielsweise Abwasser zur Kläranlage gepumpt wird oder nicht, wird nur durch den Parameter Füllstand im Pumpwerksbecken entschieden, egal ob die Kläranlage das Abwasser braucht oder nicht. Nach der Modernisierung der Außenbauwerke und datentechnischen Verknüpfung mit der Kläranlage Heringhausen wird in der zentralen Kläranlagensteuerung entschieden, ob die Pumpen fördern oder noch warten müssen. Diese neuen Prozessbedingungen sind ebenfalls als weitere Lastfälle zu erstellen und das CBR-Modell ist entsprechend anzupassen.

Der bisher entwickelte CBR-Ansatz wird zudem in weiteren Vorhaben wie z.B. im aktuellen BMBF-Forschungsprojekt „InSchuKa4.0“ (BMBF, 2022) aufgegriffen und weiter entwickelt. Hier entwickeln die Autoren in Kooperation mit der Stadt Jena und verschiedenen Industriepartnern in einem großen Kanalabschnitt eine weiterentwickelte CBR-Lösung, die spezielle technische Ausrüstungen ebenfalls fallbasiert steuern werden, um ebenfalls eine bessere Reinigungsleistung der Kläranlage zu erreichen. Darüber hinaus ist auch

die Vermeidung von Abwasseraustritt aus dem Kanal in die Gewässer angestrebt. Weiterhin sollen die technischen Ausrüstungen durch spezielle Einstellungen dafür sorgen, dass das im Kanal befindliche Abwasser in seiner Strömung so beeinflusst werden kann, dass gezielt Ablagerungen von der Kanalsohle gelöst und zur Kläranlage transportiert werden können. In einem weiteren Schritt wird zudem untersucht, ob Kanalspeicherräume in Trockenperioden nach einem kurzen Starkregenfall Regenwasser aufnehmen und für Bewässerungszwecke speichern können. Diese multifunktionalen Anwendungsziele unterliegen in der Erfüllung ebenfalls unterschiedlichen Rahmenbedingungen, die wie in Heringhausen als Fälle beschrieben und mit SIMBA# modelliert werden sollen. Eine Steuerung der technischen Ausrüstungen über eingesetzte Sensoren ist denkbar, setzt aber eine Vielzahl an Echtzeitsensoren voraus, die eine hohe Investition erfordert. Der Einsatz eines CBR-Modells ist annähernd genau genug und stellt ein weitaus besseres Kosten-Nutzen-Verhältnis dar.

Je komplexer allerdings die Anwendungsziele sind und damit eine größere Datendichte erforderlich wird, desto größer wird die Gefahr von fehlerhaften Daten, wodurch unvollständige Lastfälle entstehen können, die dann nicht mehr zuverlässig eingestuft werden. Hiervon ist im Projekt „InSchuKa4.0“ auszugehen. Es wird angenommen, dass das einfache und robuste CBR-Verfahren, wie es in Heringhausen entwickelt wurde und erprobt wird, in dieser Form für „InSchuKa4.0“ nicht optimal geeignet sein wird. Um diese Problemstellung zu lösen, soll auf das sogenannte Multiple Retrieval Case-based Reasoning (MRCBR) Framework für unvollständige Datenbanken zurückgegriffen werden, da es ein statistisch genaueres Ranking für ähnliche Fälle liefert. Es vereint die Vorteile von multipler Imputation (Vervollständigung unvollständiger Daten durch statistische Bearbeitung vorhandener Daten) und CBR, wobei sowohl die Datenverteilung als auch die Datenbankstruktur erhalten bleiben. MRCBR wurde als verallgemeinertes CBR-System entwickelt und für die medizinische Entscheidungsunterstützung optimiert und getestet, kann aber auch auf andere CBR-Anforderungen erweitert werden (Löw, Hesser, & Blessing, 2019).

Es eignet sich für numerische und kategorische Variablen und für alle Arten von Missingness-Bedingungen. Um ein korrektes und genaues Ranking zu gewährleisten, wenn Variablenwerte im Datensatz fehlen, haben (Löw, Hesser, & Blessing, 2019) einen Mechanismus entwickelt, der die Stärken der multiplen Imputation und des fallbasierten Reasonings kombiniert. Dieser Ansatz erzeugt aus dem unvollständigen Datensatz mehrere vollständige Imputationsdatensätze und führt die Suche mit eigenen globalen Ähnlichkeiten für jeden dieser Datensätze durch. Auf diese Weise spiegelt jeder Abfragesatz die Unsicherheit über das wahre Ergebnis aufgrund der Imputation wider. Die globalen Ähnlichkeiten dieser verschiedenen Abfragesätze werden miteinander verschmolzen, um das endgültige Gesamtergebnis zu erhalten. Dieser gepoolte Abrufmechanismus liefert ein besseres statistisches Ergebnis, da er die mögliche Verteilung der imputierten fehlenden

Werte jeder Variablen und ihre Auswirkungen auf das CBR-Ähnlichkeitsverhalten berücksichtigt.

Die bisher gewonnenen Erkenntnisse bestätigen bereits die Richtigkeit der Projektidee, durch den Einsatz von computer-gestützter Modellierung, KI-basierter Entscheidungsunterstützung miot dem CBR-Verfahren sowie moderner Mess- und Automationstechnik den Betrieb kleine und mittlere Abwasseranlagen wie in der Gemeinde Diemelsee flexibel an die jeweiligen Anforderungen anzupassen und somit eine sensible Umwelt in energieeffizienter Form signifikant zu schützen.

Autoren:

Prof. Günter Müller-Czygan
Hochschule Hof iwe Institut für Wasser- und Energiemanagement
Alfons-Goppel-Platz 1, 95028 Hof
E-Mail: guenter.mueller-czygan@hof-university.de

Prof. Dr.-Ing. Jürgen Wiese
Hochschule Magdeburg-Stendal
Siedlungswasserwirtschaft, Schwerpunkt Abwasser
Breitscheidstraße 2, 39114 Magdeburg
E-Mail: juergen.wiese@h2.de

Viktoriya Tarasyuk, PhD
Hochschule Hof iwe Institut für Wasser- und Energiemanagement
Alfons-Goppel-Platz 1, 95028 Hof
E-Mail: viktoriya.tarasyuk@hof-university.de

Dr.-Ing. Ralf Tscherpetzki
Hochschule Magdeburg-Stendal
Siedlungswasserwirtschaft, Schwerpunkt Abwasser
Breitscheidstraße 2, 39114 Magdeburg
E-Mail: ralf.tscherpetzki@h2.de

Literaturverzeichnis

Alex, J., & Ogurek, M. (2020). Simulation als Tool zur Quantifizierung von Kapazitätsreserven. In *Wiener Mitteilungen* (S. 147-166). Wien: TU Wien.

Atanassov, A., & Boshnakov, K. (2014). Case-Based Reasoning Control of Biological Wastewater Treatment. *ournal of Chemical Technology and Metallurgy*, S. 503-510.

BMBF. (2022). *InSchuKa 4.0 – Kombinierter Infrastruktur- und Umwelt-Schutz durch KI-basierte Kanalnetzbewirtschaftung*. Von https://www.bmbf-wax.de/verbundvorhaben/inschuka4-0/ abgerufen

Chabroswki, K. (2022). *Überlegungen zum Einsatz von Case-Based Reasoning zur Erkennung der Zulaufbelastung einer SBR-Kläranlage.* Magdeburg: Bachelorarbeit an der Hochschule Magdeburg-Stendal (unveröffentlicht).

DFKI. (2022). *Online-Präsenz für das CBR Tool my CBR*. Von http://bycbr-project.org abgerufen

Gino Sophia, S. G., Ceronmani Sharmila, V., Suchitra, S., Sudalai Muthu, T., & Pavithra, B. (2020). Water management using genetic algorithm-based machine learning. *Soft computing*, S. 17153-17165.

IFAK. (2022). *SIMBA#.* Von https://www.ifak.eu/de/produkte/simba abgerufen

Koch, G., Kühni, M., Gujer, W., & Siegrist, H. (2000). Calibration and validation of Activated Sludge Model No. 3 for Swiss municipal wastewater. *Water Research*, S. 3580–3590.

Krieg, H. J. (2000). uswirkungen von Mischwasserimmissionen auf die Mikroorganismen der heterotrophen Aufwuchsgemeinschaft des Isebekkanals (Hamburg). *Limnologica,*, S. 144-155.

Krovvidy, S., & Wee, W. G. (1993). Wastewater treatment systems from case-based reasoning. In *Case-Based Learning.* Bosten: Springer.

Löw, N., Hesser, J., & Blessing, M. (2019). Multiple retrieval case-based reasoning for incomplete datasets. *Journal of biomedical informatics,*.

Mikkelsen, P. (2018). Water solutions for smart cities – Smart water systems make cities more liveable, resilient and sustainable. In P. Mikkelsen, *Smart Cities - creating liveable, sustainable and prosperous societies.* State of Green.

Müller-Czygan, G. (2020). Von vernetzten Kläranlagen bis zur Vorhersage von Gullyüberflutungen. *wwt*, 31-34.

Müller-Czygan, G. (001-007. 12 2021). HELIP and “Anyway Strategy” - Two Human-Centred Methods for Successful Digitalization in the Water Industry. *Annals of Social Sciences & Management Studies*.

Müller-Czygan, G. (2021). Industrie 4.0 für die Wasserwirtschaft. In P. Dortans (Hrsg.), *Innovation Race.* Berlin.

Müller-Czygan, G. (2022). Erfolgsfaktoren für die Digitalisierung in der Wasserwirtschaft. *Band 174 "Wasserhygientage 02.- 03. Februar 2022" der Schriftenreihe des Vereins WaBoLu e.V.* Dessau: WaBoLu e.V.

Müller-Czygan, G., Wiese, J., & Vogel, J. (2021). Fallbasierte Steuerungen – ein Zukunftsmodell für mehr Betriebsffizienz und Anlagensicherheit in der Abwassertechnik!? *Korrespondenz Abwasser*, S. 102-106.

Müller-Czygan, G., Wimmer, M., Wagner, C., & Tarasyuk, V. (2021). How Does Digitization Succeed in the Municipal Water Sector? The WaterExe4.0 Meta-Study Identifies Barriers as well as Success Factors, and Reveals Expectations for the Future. *Energis*.

Quan, L. Y. (2018). Prediction of effluent Ammonia Nitrogen using FNN-based CBR. *37th Chinese Control Conference (CCC)*, S. 6137-6142.

Rahim, M. S., Nguyen, K. A., Stewart, R. A., Giurco, D., & Blumenstein, M. (2020). Machine learning and data analytic techniques in digital water metering: A review. *Water*.

Sarabia, Y. R., Bermúdez, X. C., Martinez, R. F., Rodríguez, Z. H., Moya, A. M., & Lorenzo, M. M. (2005). CBR-ANN hybrid model to optimize the sequence of wastewater treatments. *ITEE*, S. 711-720.

Schmitt, S., Stahl, A., & Hansen, J. (2002). A CBR Approach to Identification and Counteraction for Harmful Microorganisms in Wastewater Treatment Plants. *Proceedings International Conference on Artificial Intelligence*.

Trstenjak, B., Palasek, B., & Trstenjak, J. (2019). A decision support system for the prediction of wastewater pumping station failures based on CBR continuous learning model. *Engineering, Technology & Applied Science Research*, S. 4745-4749.

Vijayaraghavan, G., & Jayalakshmi, M. (2015). A quick review on applications of fuzzy logic in waste water treatment. *Int. J. Res. Appl. Sci. Eng. Technol*, S. 421-425.

Wiese, J. (2004). *Entwicklung von Strategien für einen integrierten Betrieb von SBR-Kläranlagen und Mischkanalisationen.* Kaiserslautern: TU Kaiserslautern.

Wiese, J. (2005). *Entwicklung von Strategien für einen integrierten Betrieb von SBR-Kläranlagen und Mischkanalisationen - Simulation und Großtechnik,.* Kaiserslautern: TU Kaiserslautern.

Wiese, J., Stahl, A., & Hansen, J. (2005). Applying and optimizing case-based reasoning for wastewater treatment systems. *AI Communications*, S. 269-279.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1473

Entscheidungsunterstützung bei der Nutzung von Kläranlagen für Agrarsysteme

Markus Junker [1], Ansgar Bernardi [2]

Abstract:

Absehbare Entwicklungen wie wachsende Bevölkerung in Städten, knappe Rohstoffe und Düngemittel, oder der Wunsch nach kurzen Transportwegen und regionalen Nahrungsmitteln motivieren neue Visionen für Agrarsysteme und Nahrungsproduktion der Zukunft. Ein innovativer Ansatz zur Nahrungsmittelproduktion besteht in der direkten Nutzung der bei einer Kläranlage anfallenden Ressourcen in einem baulich angegliederten Agrarsystem. Die Nutzung von Nährstoffen in den Stoffströmen der Kläranlage geht mit einigen Herausforderungen einher. Im Projekt SUSKULT[3] wird untersucht, wie eine auf die Bedürfnisse solcher Agrarsysteme ausgerichtete Kläranlage gestaltet sein muss. In unserer Arbeit haben wir den Einsatz von Entscheidungsunterstützung in Form Maschineller Lernverfahren für die Vorhersage von Stoffkonzentrationen sowie Optimierungsverfahren für die Zusammenstellung von Nährlösungen aus verschiedenen Stoffströmen qualitativ untersucht. Es hat sich gezeigt, dass die entsprechenden Verfahren zumindest theoretisch an verschiedenen Stellen einen Beitrag liefern können.

Keywords: Kläranlagen; Agrarsysteme; Nahrungsmittelproduktion; Entscheidungsunterstützung; LSTM; SVM; Interior Point Algorithmus

1 Motivation

Absehbare Entwicklungen wie wachsende Bevölkerung in Städten, knappe Rohstoffe und Düngemittel, oder der Wunsch nach kurzen Transportwegen und regionalen Nahrungsmitteln motivieren neue Visionen für Agrarsysteme und Nahrungsproduktion der Zukunft. Das Projekt SUSKULT untersucht einen innovativen Ansatz zur stadtnahen Nahrungsmittelproduktion durch Einsatz von Hydroponik, bei dem die Pflanzen im Rahmen einer Indoor-Kultivierung boden- bzw. erdelos unter Einsatz mineralischer Nährstofflösungen wachsen und gedeihen. Abbildung 1 zeigt das in SUSKULT geplante Zusammenspiel von Abwassertechnik und Agrarsystemen im Schaubild: Die Abwassertechnik versorgt das

[1] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Trippstadter Str. 122, 67663 Kaiserslautern, Germany markus.junker@dfki.de

[2] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Trippstadter Str. 122, 67663 Kaiserslautern, Germany ansgar.bernardi@dfki.de

[3] Das Projekt „SUSKULT – Entwicklung eines nachhaltigen Kultivierungssystems für Nahrungsmittel resilienter Metropolregionen“ (FKZ 031B0728) wird im Rahmen der Fördermaßnahme „Agrarsysteme der Zukunft“ im Rahmen der »Nationalen Forschungsstrategie „BioÖkonomie 2030“ der deutschen Bundesregierung durch das Bundesministerium für Bildung und Forschung (BMBF) gefördert.

doi:10.18420/inf2022_126

Agrarsystem mit Wärme, Wasser, Kohlendioxid sowie zum Pflanzenwachstum notwendigen Nährstoffen.

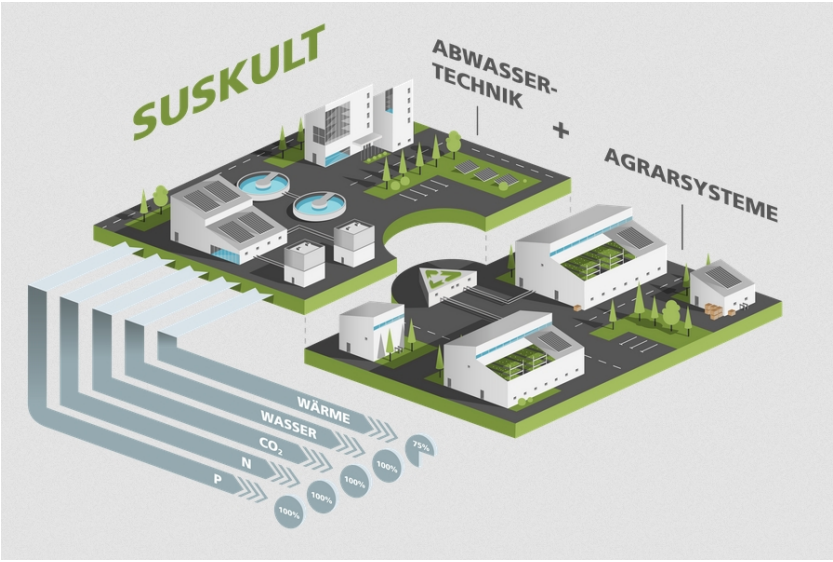


Abb. 1: Nutzung von Abwasser für Agrarsysteme in SUSKULT (© Fraunhofer UMSICHT)

Zur Nutzung der Kläranlage für die Nahrungsmittelproduktion müssen ausgewählte Stoffströme bestmöglich für die Nutzung durch die Pflanzen ausgewählt, gesteuert und kombiniert werden. Die Stoffkonzentrationen in Strömen einer Kläranlage hängen von verschiedenen Faktoren ab. Entscheidend ist zunächst der Zufluss zu der Kläranlage, der in seinem Umfang und seiner Zusammensetzung nach nicht beeinflussbar ist. Er hängt von Tageszeiten, Wochentagen, der Jahreszeit, aber auch Wetterlagen und auch von Einzeleinleitungen durch Industrie und Landwirtschaft ab. Die Stoffkonzentrationen in einer Kläranlage können in gewissem Umfang durch Prozessparameter gesteuert werden. Die Seite der Nahrungsversorgung für die Pflanzen ist dadurch gekennzeichnet, dass verschiedene Pflanzenarten in verschiedenen Wachstumsstadien individuelle Anforderungen an die Nährlösung hinsichtlich Nährstoffbedarfen haben. Eine besondere Rolle spielen dabei auch in den Stoffströmen eventuell noch vorhandene Schadstoffe und die Toleranzgrenzen der Pflanzen bei diesen Schadstoffen (beispielsweise Streusalz im Winter).

In der vorliegenden Arbeit wurden zwei Ansätze betrachtet, die einen Betrag zur Optimierung der Pflanzenversorgung mit Hilfe von Stoffströmen aus der Kläranlage leisten können:

- Die **Vorhersage von Stoffkonzentrationen** mit dem Ziel einer prädiktiven Steuerung der Kläranlage im Sinne der für die Pflanzenproduktion geeigneten Konzentrationen.

- Die **Mischung von verschiedenen Stoffströmen** mit jeweils individuellen Stoffkonzentrationen zur Erzeugung von für die Pflanzenproduktion geeigneten Konzentrationen unter Minimierung von Nährstoffzukäufen.

Der Einsatz von KI zur Steuerung einer Kläranlage ist kein sehr aktives Forschungsfeld. Der Einsatz von KI zur Nutzung von Stoffströmen einer Kläranlage für Agrarsysteme wurde unseres Wissens nach bisher nicht betrachtet. Einige bisherige Arbeiten im Kontext KI und Kläranlagen seien nachfolgend exemplarisch aufgelistet: In [Ha97] wurde der von Fuzzy-Logik zur Minimierung von Stickstoff, Phosphor und Betriebskosten adressiert. [Sc04] verfolgte ebenfalls die Minimierung von Stickstoff mit Fuzzy-Logik. [Gu15] entwickelte ein Warnsystem für die Stickstoffbelastung mit Hilfe von Support Vector Machines und Neuronalen Netzen. [Eb04] hatten die Vorhersage von NH4-N am Ablauf des Belebungsbeckens zur Steuerung des Sauerstoffeintrags über ein Neuronales Netz und Fuzzy Regler zum Ziel. [Gr17] setzte Support Vector Machines und Entscheidungsbäume zur Vorhersage von Abwasserqualitätsindikatoren auf Basis der Beschreibung des Einzugsgebietes ein. In deutschen Verbundprojekt E-KLÄR [IfS] wurde unter anderem an Systemen zur Entscheidungsunterstützung bei der Auslegung von Kläranlagen durch in Bayesschen Netzen modellierte Kausalbeziehungen gearbeitet. [Ol18] beschäftigte sich mit Kostenoptimierung einer Kläranlage mit Reinforcement Learning unter Einhaltung von gesetzlichen Reinigungsvorgaben.

Heutige Kläranlagen verfügen in der Regel weder über umfangreiche, kleinräumig und kurzfristig messende Sensorik, noch bieten sie detaillierte, auf einzelne Nähr- und Schadstoffe ausgerichtete Eingriffs- und Steuerungsmöglichkeiten. Andererseits sind die optimale Nährstoffversorgung zahlreicher Nutzpflanzen im hydroponischen Anbau in ihrem zeitlichen Profil sowie der Umgang mit unerwünschten Begleitstoffen noch aktuelle Forschungsgegenstände und in vielen Details noch unbekannt. Die hier beschriebenen Arbeiten nutzen bereits vorhandene Aufzeichnungen von kooperierenden Kläranlagen aus den letzten Jahren, die jedoch in Auflösung und erfassten Stoffkonzentrationen eingeschränkt waren. Die Arbeiten mussten sich daher darauf beschränken, das Potential von KI-Ansätzen in diesem Anwendungskontext initial und zunächst rein qualitativ zu erheben und anhand der wenigen verfügbaren Daten auf Machbarkeit zu prüfen. Daher wird auf die Angabe quantitativer Ergebnisse im folgenden bewusst verzichtet.

2 Vorhersage von Stoffkonzentrationen an der Kläranlage

Durch die Abschätzung zukünftig zu erwartender Konzentrationen bestimmter Stoffe kann die Kläranlage prädiktiv im Sinne einer optimalen Aufbereitung der Stoffströme für die Pflanzenproduktion gesteuert werden. An welchen Stellen die Kläranlage konkret und durch Voraussagen mit bestimmtem zeitlichen Horizont und bestimmten Genauigkeiten tatsächlich im Sinne der Pflanzenproduktion gesteuert werden kann, ist aktuell noch Gegenstand von Forschungsarbeiten weiterer Konsortialpartner in SUSKULT.

Ziel der in diesem Abschnitt beschriebenen Arbeiten war es zunächst, Stoffkonzentrationen in einem Zeitfenster von wenigen Stunden vorherzusagen sowie das Potential der zusätzlichen Kenntnis/Prognose von Zuflussmengen zu bewerten. In einem weiteren Teil der Arbeiten wurde versucht, das Über-/Unterschreiten vorgegebener Schwellwerte für Stoffkonzentrationen am Folgetag vorherzusagen.

2.1 Vorhersage im Stundenbereich

Die ersten Machbarkseitsanalyse hatten den Einsatz eines Lernverfahrens zur Vorhersage von Stoffkonzentrationen im Bereich von einigen Stunden zum Ziel. Grundlage war ein an der Kläranlage Bochum-Ölbachtal erhobener Datensatz, welcher über den Zeitraum 1.1.2015 bis 31.12.2018 stundengenau die Messwerte für Stickstoff (N) und Phosphor (P) enthielt. Insgesamt erhielt der Datensatz 11.736 Messzeitpunkte für die entsprechenden Werte. Aufgrund des ausreichend erscheinenden Umfangs des Datensatzes entschieden wir uns für den Einsatz eines neuronalen Netzes zur Vorhersage von Messwerten in einigen Stunden, basierend auf der Historie der Messwerte. Konkret wurde ein LSTM Netz (Long Short-Term Memory, [HS97]) mit in Summe 4x200 Zellen eingesetzt. Besonders attraktiv an diesen Netzen erschien, dass sie theoretisch in der Lage sind, Zyklen beliebiger Länge in den Daten zu erkennen und nutzbar zu machen (insbesondere Tagesläufe, Wochenläufe, jahreszeitliche Regelmäßigkeiten). Weiterhin sind sie theoretisch in der Lage, besondere Zulaufereignisse (z.B. Regenereignisse, unregelmäßig auftretende Industriezuläufe) frühzeitig anhand von zeitlichen Mustern zu erkennen. Es wurde davon ausgegangen, dass die durch LSTM Netze mögliche Erkennung von Zyklen und Zulaufereignissen zur Erhöhung der Vorhersagegenauigkeit beitragen kann. Um die mögliche Vorhersagegenauigkeit weiter zu erhöhen, wurden dem LSTM Netz auch Informationen zum Wochentag und Monat der Messung zur Verfügung gestellt.

2.1.1 Initiale Erkenntnisse

In initialen Experimenten zeigte sich, dass das LSTM-Netz im Bereich bis zu 3 Stunden relativ genaue Vorhersagen leisten kann (vgl. Abbildung 2). Die Vorhersage der Messwerte über 3 Stunden hinaus zeigte zunehmend größere Abweichungen von der Realität. Besonders große Abweichungen von den tatsächlich gemessenen Werten wurde zu Zeiten beobachtet, in denen der Zufluss zur Kläranlage erhöht war, vermutlich in vielen Fällen durch Regenereignisse.

2.1.2 Berücksichtigung von Zuflussmengen

Es stellte sich damit die Frage, ob die Genauigkeit der Vorhersage einer erwarteten Stoffkonzentration (z.B. in drei Stunden) durch die Kenntnis der zukünftigen Durchflussmenge (z.B. der Durchflussmengen in einer, zwei und in drei Stunden) erhöht werden kann. In einer

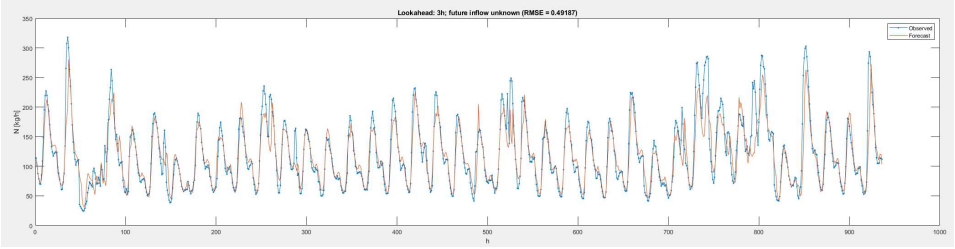


Abb. 2: Vorhersage der Stickstoffkonzentration in 3 Stunden (rot) im Vergleich zu Messwert in drei Stunden (blau)

echten Anwendung können die zukünftigen Durchflussmengen lediglich abgeschätzt werden. Die hier verwendeten, aufgezeichneten und retrospektiv betrachteten Datensätze enthielten jedoch zu jedem Messzeitpunkt auch die tatsächliche Durchflussmenge an der Kläranlage. Es war daher möglich, ausgehend von einem gewählten Startzeitpunkt den Einfluss der zukünftigen tatsächlichen Durchflussmengen auf die Vorhersagegenauigkeit zu überprüfen und so zu zeigen, was mit einer sehr guten Abschätzung zukünftiger Durchflussmengen an Verbesserung der Vorhersage von Stoffkonzentrationen erzielbar wäre.

Die entsprechenden Experimente zeigten, dass die Kenntnis der zukünftigen Durchflussmengen bis zum Vorhersagezeitpunkt die Genauigkeit in der Tat wesentlich erhöhen kann. Mit Kenntnis der stündlichen Durchflussmengen in den nächsten 6 Stunden war auch eine deutlich zuverlässigere Vorhersage der Messwerte in 6 Stunden möglich.

Im Datensatz waren zu jedem Zeitpunkt auch die Durchflussmengen angegeben. Insofern war es kein Problem, den exakten Zufluss zum Vorhersagezeitpunkt mit in die Vorhersage eingehen zu lassen. Unter realen Bedingungen ist der zukünftige Zufluss nicht bekannt. Ersatzweise besteht die Möglichkeit, vorgelagerte Messungen im verzweigten Zulauf der Kläranlage vorzunehmen (die sich dann erst mit einiger Verzögerung bei der Kläranlage manifestieren), Niederschläge im Einzugsgebiet zu messen oder Niederschlagsvorhersagen zu nutzen – mit zunehmendem Vorhersagehorizont und zunehmender Ungenauigkeit im Hinblick auf die tatsächlichen zukünftigen Zuläufe.

Für eine experimentelle Evaluierung wurden Niederschläge im Umfeld der Kläranlage herangezogen, genauer, Wetterdaten, die am etwa 40km von der Kläranlage Bochum-Ölbachtal des Ruhrverbandes entfernten Flughafen Düsseldorf erhoben wurden. Zum Zeitpunkt der Durchführung der Arbeiten war die Verfügbarkeit näherer Wettermessstationen nicht bekannt. Eine erste Überprüfung zeigte, dass Niederschlagsereignisse am Flughafen mit etwa zwei Stunden Verzögerung im Zufluss der Kläranlage sichtbar werden. Die Ergebnisse aus den Experimenten zeigten weiter, dass die Genauigkeit durch die Kenntnis der verwendeten Wetterdaten statt der tatsächlichen zukünftigen Zuflüsse nur marginal im Vergleich zu den Vorhersagen ohne jegliche Zuflussinformationen erhöht wird. Mögliche Ursachen hierfür sind die Wahl des Standorts der Messung, die die tatsächlichen Regenereignisse im

Einzugsbereich der Kläranlage möglicherweise nur unzureichend wiedergeben, sowie der zu geringe Vorhersagezeitraum für die erhöhten Zuläufe von nur ca. zwei Stunden.

2.1.3 Lernrate

In einem weiteren Schritt wurde untersucht, wie lange der Beobachtungszeitraum sein muss, um eine gute Vorhersagegenauigkeit der Messwerte in 3 Stunden zu erreichen. Es zeigte sich hierbei, dass erste „Lernerfolge“ bereits nach einer Woche sichtbar werden, dass das Verfahren aber deutlich von der Verfügbarkeit wesentlich längerer Beobachtungszeiträume profitiert. Auch bei der Verlängerung des Bobachtungszeitraums über ein Jahr hinaus wurden weitere Genauigkeitsverbesserungen beobachtet. Es ist davon auszugehen, dass bei Verfügbarkeit von noch längeren Messreihen noch größere Genauigkeiten erreicht werden können.

2.1.4 Schätzwerte für Sensordaten

Es ist zu erwarten, dass viele für die Pflanzenproduktion voraussichtlich notwendige Messungen von Stoffkonzentrationen, insbesondere auch von Schadstoffen, aufwändig und teuer sind. Ein weitere interessante Anwendung für Lernverfahren besteht damit in der Abschätzung aktueller, nicht gemessenen / nicht kostengünstig messbarer Stoffkonzentrationen aus der Historie und aus dem Vorliegen aktueller anderer Messwerte. Entsprechende Experimente für die Vorhersage von Stickstoff und Phosphor zeigten vielversprechende Resultate. So war aus der Kenntnis der aktuellen Konzentrationen von Stickstoffhalt, dem chemische Sauerstoffbedarf (CSB) und der Zuflussmenge die aktuelle Konzentration von Phosphor sehr präzise vorhersagbar.

2.2 Vorhersage von Tagesdurchschnittswerten

Auf Kläranlagen wird die Durchflussmenge zwar kontinuierlich aufgezeichnet, die Stoffkonzentrationen im Zulauf werden jedoch aktuell in der Regel nur in Tagesdurchschnitten und diskontinuierlich gemessen. Vor diesem Hintergrund hatte die zweite Machbarkseitsanalyse den Einsatz eines Lernverfahrens zur Vorhersage des Tagesdurchschnittes am Folgetag zum Ziel. Der den diesen Untersuchungen zu Grunde liegende Datensatz enthielt in täglicher Auflösung für den Zeitraum 10.01.2016 bis 21.12.2019 jeweils von Montag bis Donnerstag Tagesdurchschnittswerte für den Belebungsbeckeneingang. Daten an den fehlenden Tagen (Freitag bis Sonntag) wurden testweise auf 0 gesetzt bzw. linear interpoliert. Auch wurden wieder Tages- und Monatsinformationen hinzugefügt. Die experimentelle Evaluierung auf Basis dieses Datenbestands zeigte bei der Verwendung von LSTM Netzen keinen Lernerfolg.

Es war zu vermuten, dass der geringe Umfang des Datenbestandes in Verbindung mit der Diskontinuität (fehlende Tagesmesswerte) Ursache für die schlechten Ergebnisse mit LSTM-Netzen ist. In einem weiteren Schritt wurde daher der Einsatz einer Support Vector Machine (SVM) als klassisches Klassifikationsverfahren angestrebt. Zur Umsetzung der Vorhersage als Klassifikationsproblem wurde ein Zeitfenster vorgesehen, das dem Verfahren neben den aktuellen Tagesmesswerte auch die Messwerte am Vortag und vor zwei Tagen zur Vorhersage bereitstellt. Weiterhin wurden als Klassen Testergebnisse für Messwerte definiert. Diese hatten die Form: „Messwert am morgigen Tag ist im unteren Drittel/mittleren Drittel/oberen Drittel der üblicherweise gemessenen Werte“. Zeitfenster, die fehlende Messwerte enthielten, wurden komplett vom Datensatz ausgeschlossen. Insgesamt konnten so aus den verfügbaren 1.471 Messtagen lediglich 176 vollständige Datensets zum Lernen und für die Evaluierung verwendet werden.

Mit der entsprechen Datengrundlage wurde geprüft, inwieweit sich mit der SVM „Detektoren“ für erhöhte bzw. verminderte Messwerte am Folgetag (d.h. Messwerte im unteren und oberen Drittel des üblichen Bereichs) entwickeln lassen. Mithilfe von Konfusionsmatrizen und Receiver Operating Characterics (ROC) Diagrammen konnte gezeigt werden, dass sich ein Lernerfolg einstellt. Allerdings kommt es sowohl bei den False Positives als auch bei den False Negatives zu signifkanten Fehlerraten (siehe Abbildung 3). Ob und bei welchen Arbeitspunkten der ROC-Kurve der Einsatz eines Lernverfahrens gegebenenfalls sinnvoll möglich ist, lässt sich erst im Rahmen einer konkreten Steuerung nach Durchführung einer Kosten/Nutzen-Rechnung ermitteln.

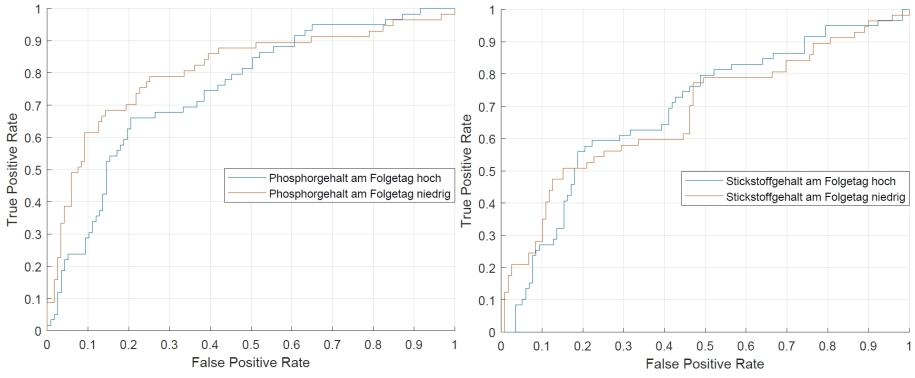


Abb. 3: Receiver Operating Characterics zur Vorhersage von Schwellwertunterschreitungen und -überschreitungen für Phosphor und Stickstoff

3 Zusammenstellung von Nährlösungen für die Pflanzenproduktion

In einem zweiten Arbeitsschwerpunkt wurde direkt die Schnittstelle zwischen der Kläranlage und der Pflanzenproduktion adressiert. Ziel war es, als gegeben angenommene, verschiedene

Stoffströme der Kläranlage bestmöglichst zur Erzeugung von Nährstofflösungen zu verwenden und Zukäufe von Nährstoffen aus Kostengründen (Geld oder entstehende „ökologische Kosten“) zu minimieren.

3.1 Modellbildung

Im ersten Schritt der Arbeiten wurden die wesentlichen Abläufe/Zusammenhänge an der Schnittstelle zwischen Kläranlage und Pflanzenproduktion modelliert (siehe Abbildung 4).

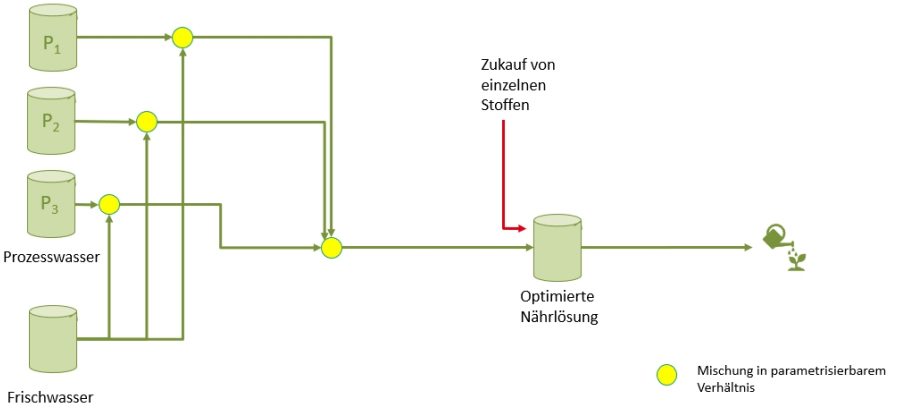


Abb. 4: Modell zur Erzeugung von Nährlösungen aus Prozesswasserchargen

Auf der Seite der Kläranlage wurde aus Vereinfachungsgründen davon ausgegangen, dass verschiedene (Prozess-)wasserchargen („Tanks“) mit verschiedenen Zusammensetzungen hinsichtlich Nähr- und Schadstoffen gegeben sind. Dem liegt die Idee zu Grunde, dass diese unterschiedlichen Wasserchargen zu verschiedenen Betriebszeiten der Kläranlage in Tanks zwischengespeichert werden für eine spätere, möglichst optimale Verwendung. Auf der Seite der Pflanzenproduktion sollte unter Verwendung der Wasserchargen eine optimale Nährlösung generiert werden. Diese zeichnet sich aus durch (i) das Bereitstellen exakt vorgegebener Nährstoffkonzentrationen in der Nährlösung sowie (ii) das Nichtüberschreiten von vorgegebenen spezifischer Grenzwerten für verschiedener Schadstoffe. Zur Erzeugung einer optimalen Nährlösung aus den gegebenen Wasserchargen wurden folgende Verarbeitungsschritte modelliert:

- Für jede Wassercharge ist eine beliebige spezifische Verdünnung mit Wasser möglich.
- Die (verdünnten) Wasserchargen können in beliebigen Verhältnissen gemischt werden.
- Gekaufte Nährstoffe können in beliebigen Konzentrationen zugegeben werden. Der Zukauf von Nährstoffen ist mit für den Nährstoff spezifischen Kosten verbunden. Diese Kosten können einem Geldbetrag entsprechen oder einem Maß für die ökologischen Kosten.

Es wäre auch möglich gewesen, das Entfernen von Schadstoffen für die Pflanze oder die Transformation von Stoffen als zulässige, möglicherweise mit weiteren Kosten verbundenen Verarbeitungsschritte vorzusehen. Da zumindest heutige Kläranlagen hierzu noch keine Möglichkeiten haben, wurde auf diese Option verzichtet.

Die Messung von Nähr- und Stoffkonzentrationen selbst ist ebenfalls mit spezifischen Kosten verbunden, die in einer praktischen Anwendung einbezogen werden müssen. Derartige Kosten sind allerdings mit der vorgestellten Modellierung nicht erfassbar und eine Erweiterung ist aktuell auch nicht geplant.

Das Optimierungsproblem wurde wie folgt formuliert: sei n die Anzahl der Prozesswassertanks, k die Anzahl der zu betrachtenden Nährstoffe und l die Anzahl der zu betrachtenden Schadstoffe. Dann bezeichnet $\overrightarrow{p}_{i,j}$ mit $i = 1..n$ und $j = 1..k+l$ die Konzentration des Stoffes j im Tank i. Weiter bezeichne $c_{1..k}$ die Kosten für die Zugabe von Nährstoffe $1..k$ und m die Menge der zu erzeugenden Nährlösung. Gesucht ist dann ein Vektor $\overrightarrow{x}$ mit $\sum_{i=1..k} x_{2n+i} c_i$ minimal, wobei $x_1..x_n$ der Entnahme aus den Prozesswassertanks $1..n$ entspricht, $x_{n+1}..x_{2n}$ dem Verdünnungsfaktor für Entnahme $1..n$ und $x_{2n+1}..x_{2n+k}$ der Zugabe von Nährstoff $1..k$. Das Optimum muss unter den folgenden Nebenbedingungen gefunden werden:

- Die Konzentrationen der Nährstoffe $1..k$ müssen den Vorgaben z_i entsprechen, d.h. $\frac{\sum_{i=1..n}(x_i p_{i,1} x_{n+i}^{-1})}{\sum_{i=1..n} x_i x_{n+i}^{-1})} + x_{2n+1} = z_1, ..., \frac{\sum_{i=1..n}(x_i p_{i,1} x_{n+i}^{-1})}{\sum_{i=1..n} x_i x_{n+i}^{-1})} + x_{2n+k} = z_k$
- Die Konzentrationen der Schadstoffe $1..l$ müssen unterhalb der Grenzwerte z_{k+i} liegen, d.h. $\frac{\sum_{i=1..n}(x_i * p_{i,1} * x_{n+i}^{-1})}{\sum_{i=1..n} x_i * x_{n+i}^{-1})} - x_{2n+k+1} < z_{k+1}, ..., \frac{\sum_{i=1..n}(x_i * p_{i,1} * x_{n+i}^{-1})}{\sum_{i=1..n} x_i * x_{n+i}^{-1})} - x_{2n+k+l} < z_{k+l}$
- Es muss die vorgegebene Menge an Nährlösung erzeugt werden: $\sum_{i=1..n}(x_i * x_{n+i}) = m$

Zur Lösung des Optimierungsproblems wurde ein Interior Point Algorithmus eingesetzt, welcher das Minimum einer beschränkten nichtlinearen multivariablen Funktion sucht [BGN00]. Ausgehend von einer (händisch) vorgegebener Startkonfiguration (z.B. der unverdünnten Verwendung des Inhalts eines Prozesswassertanks ohne Zukäufe) findet der Algorithmus ein lokales Optimum hinsichtlich der Kosten. Ein globales Optimum kann nicht garantiert werden. Allerdings ist durch die Vorgehensweise garantiert, dass die vom Algorithmus gefundene Lösung in jedem Fall mindestens so kostengünstig ist, wie die händisch vorgegebene Startkonfiguration.

3.2 Experimentelle Evaluierung

Die Modellierung wurde an einem vereinfachten Szenarien getestet anhand dreier Prozesswassertanks P_1, P_2 und P_3 mit einer Kombinationen dreier Nährstoffe (Stickstoff (N), Phosphor (P) und Kalium (K)) und einem nicht näher definieren Schadstoff S. Die Kosten für die Zugabe von N, P oder K wurden als gleich angesehen. Die konkreten

Nährstoff- und Schadstoffkonzentrationen von P_1, P_2 und P_3 und die Anforderungen an die optimale Nährlösung wurden in den Beispielen künstlich gesetzt, um die Plausibilität der Optimierungsergebnisse in Grenzfällen untersuchen zu können. Beispielsweise wurden die Zusammensetzungen von P_1, P_2 und P_3 so gewählt, dass jeder der Tanks jeweils einen Nährstoff enthält, aber zusätzlich P_3 noch einen Schadstoff in kritischer Konzentration. Als Startpunkt für die Optimierung wurde jeweils die Mischung von P_1, P_2 und P_3 in gleichen Verhältnissen, ohne Verdünnungen und ohne Zukauf von Rohstoffen gewählt. Ziel war jeweils die Erzeugung von 100 Litern Nährlösung. Anhand der Beispiele konnte das erwartungsgemäße Verhalten des Modells und Optimierungsverfahrens verifiziert werden.

Im nächsten Schritt wurden dann tatsächliche Stoffkonzentrationen aus Prozesswasserproben von Kläranlagen zu verschiedenen Zeitpunkten herangezogen in Kombination mit Vorgaben für die optimale Nährlösung von Tomaten. Es zeigte sich, dass die optimale Lösung des Algorithmus in der Regel lediglich eine der drei Wasserchargen heranzog und diese nach Verdünnung und Zukauf von Nährstoffen zur optimalen Nährlösung wurde. Insbesondere kam es in den Beispielen mit realen Stoffkombinationen nicht zu einer Mischung von Chargen aus verschiedenen Tanks.

Die Ergebnisse zeigen, dass mit Hilfe des Algorithmus verschiedene Wasserchargen aus der Kläranlage kostenreduzierend verdünnt, kombiniert und nach Zukauf von Rohstoffen als Nährlösung in der Pflanzenproduktion eingesetzt werden können. Die sich in den gewählten realen Wasserchargen findenden Stoffkonzentrationen waren jedoch zu ähnlich, um durch ihre Kombination zu Einsparungen zu führen.

4 Zusammenfassung und Ausblick

In der vorliegenden Arbeit wurden die potenziellen Einsatzmöglichkeiten von Entscheidungsunterstützung bei der Nutzung von Prozessströmen einer Kläranlage für Agrarsysteme qualitativ untersucht. Die Analysen zeigten, dass Vorhersagen zu Stoffkonzentration, wie sie zur prädiktiven Steuerung der Kläranlage im Sinne der Agrarsystemen eingesetzt werden kann, prinzipiell möglich ist. Die entsprechenden experimentellen Evaluierungen konnten aus Mangel entsprechender Messwerte lediglich anhand der Vorhersage zukünftiger Stickstoff und Phorsphorkonzentrationen erfolgen und zeigte die prinzipiellen Grenzen der Vorhersagehorizonte für die vorhandenen historischen Daten und speziell diese Stoffkonzentrationen. Die Analysen zeigten auch, dass zur Erhöhung von Genauigkeiten und Vorhersagehorizonte die Kenntnis von zukünftigen Zuflussmengen einen erheblichen Beitrag leisten kann. Dies muss gegebenenfalls noch genauer untersucht werden. Inwieweit bei anderen Nährstoffen und Schadstoffen bessere oder schlechter Genauigkeiten zu erzielen sind und in welchem Umfang die erzielbaren Vorhersagegenauigkeiten in SUSKULT zur Steuerung der Kläranlage tatsächlich genutzt werden können ist noch offen.

Der zweite Schwerpunkt der Arbeiten betraf die optimale Auswahl und Kombination von Prozessströmen im Sinne der Agrarsysteme. Die Analysen zeigten, dass es über

Optimierungsansätze potenziell möglich ist, die monetären oder ökologischen Kosten bei gegebenen Prozessströmen weiter zu reduzieren. Entscheidend zur Ausschöpfung dieses Potentials wäre die Steuerung der Kläranlage, damit sie von ihrer Nähr- und Schadstoffkonzentrationen her komplementäre Prozessströme generiert.

Die hier beschriebenen Arbeiten wurden auf relativ eingeschränkten Datensätzen und konnten nur einen qualitativen Eindruck davon vermitteln, wo Entscheidungsunterstützung in SUSKULT möglicherweise eingesetzt werden kann. Zentral für weiterführende Arbeiten ist der Abgleich der Ergebnisse mit neuen Messdaten aus den Kläranlagen, die sowohl umfangreicher sind als auch bisher nicht erhobene und für den Pflanzenbau relevante Schadstoffkonzentrationen enthalten.

Literaturverzeichnis

[BGN00] Byrd, Richard; Gilbert, Jean Charles; Nocedal, Jorge: A Trust Region Method Based on Interior Point Techniques for Nonlinear Programming. Mathematical Programming, 89:149–185, 11 2000.

[Eb04] Ebel, Alexander; Bongards, Michael; Westenberger, Hartmut; Graner, Martin: Prädiktive Regelung von Kläranlagen mit neuronalen Netzen. Wasser und Abfall, 51, 01 2004.

[Gr17] Granata, Francesco; Papirio, Stefano; Esposito, Giovanni; Gargano, Rudy; De Marinis, Giovanni: Machine Learning Algorithms for the Forecasting of Wastewater Quality Indicators. Water, 9(2), 2017.

[Gu15] Guo, Hong; Jeong, Kwanho; Lim, Ji Young; Jo, J.; Kim, Young Mo; Park, Jongkwan; Kim, Joon; Cho, Kyung Hwa: Prediction of effluent concentration in a wastewater treatment plant using machine learning models. Journal of environmental sciences, 32:90–101, 2015.

[Ha97] Hansen, Joachim: Der Einsatz von Fuzzy Control für Regelungsaufgaben im Bereich der Nährstoffelimination in kommunalen Kläranlagen. Dissertation, Universität Kaiserslautern, 1997.

[HS97] Hochreiter, Sepp; Schmidhuber, Jürgen: Long Short-Term Memory. Neural Computation, 9(8):1735–1780, 11 1997.

[IfS] Institut für Siedlungswasserwirtschaft, RWTH Aachen University: E-Klaer: Entwicklung und Integration innovativer Klaeranlagentechnologien fuer den Transformationsprozess in Richtung Technikwende.

[Ol18] del Olmo, Félix Hernández; Gaudioso, Elena; Dormido, Raquel; Duro, Natividad: Tackling the start-up of a reinforcement learning agent for the control of wastewater treatment plants. Knowledge-Based Systems, 144:9–15, 2018.

[Sc04] Schönberger, Dirk: Regelungstechnische und verfahrenstechnische Maßnahmen zur Leistungssteigerung kommunaler Abwasserreinigungsanlagen. Dissertation, Universität Siegen, 2004.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1485

Requirements for a Digital Product Passport to Boost the Circular Economy

Christiane Plociennik[1], Monireh Pourjafarian[1], Shehab Saleh[2], Tabea Hagedorn[3], Alice do Carmo Precci Lopes [3], Malte Vogelgesang[3], Julian Baehr[3], Bernd Kellerer[4], Maike Jansen[5], Holger Berg[5], Martin Ruskowski[1], Liselotte Schebek[3], Andreas Ciroth[2]

Abstract: The Digital Product Passport (DPP) is a concept to collect and share product-related information along a product's lifecycle. The aim is to provide all stakeholders during the product lifecycle with the information they need such that a successful Circular Economy can be implemented. At the moment, several varieties of DPPs are being developed, most of them sector-specific. As of today, however, there is no common standard as to what a DPP should look like. This paper collects requirements for the DPP from different perspectives that must be fulfilled to make it broadly applicable.

Keywords: Digital Product Passport; Circular Economy

1 Introduction and Motivation

The Circular Economy (CE) tries to counter the over-exploitation of resources by keeping materials in the loop for as long as possible by reusing, repairing, refurbishing, and recycling products or their components, with the ultimate goal of our living within earth's ecological limits [1]. The CE is promoted by the European Green Deal [2] and the Circular Economy Action Plan [3] in the EU and via the Digital Policy Agenda for the Environment [4] by the German Federal Environmental Ministry at the German national level. An important means to facilitate the development of a true CE is the concept of the Digital Product Passport (DPP) [5]. Its aim is to create transparency along a product's lifecycle and to enable the collection and sharing of product-related data among various stakeholders. The DPP allows to exchange information such as material flows, instructions for disassembly, or whether a product contains hazardous materials. Thus, it can inform recyclers which materials or compounds to expect, enable feedback to producers about the recyclability of their products, and educate customers about the environmental impact of their purchases, to name just a few applications. Furthermore, value-added services and applications can be built on top of the information contained in the DPP, such as AI-based lifecycle assessment (LCA) [6].

[1] German Research Center for Artificial Intelligence (DFKI), Trippstadter Str. 122, 67663 Kaiserslautern, Germany
[2] GreenDelta GmbH, Kaiserdamm 13, 14057 Berlin, Germany
[3] Technische Universität Darmstadt, Karolinenplatz 5, 64289 Darmstadt, Germany
[4] CIRECON, Jetsam Service Management GmbH, Dr.-Leo-Ritter-Straße 4, 93049 Regensburg, Germany
[5] Wuppertal Institute, Doeppersberg 19, Wuppertal, Germany

doi:10.18420/inf2022_127

This can benefit the field of environmental informatics and enable new business models. Several approaches to the DPP are recently being conceptualized, implemented, and tested. A systematic requirements analysis for the DPP, however, has not yet been published. Our paper aims to close this gap. Aligning the requirements of different stakeholders and sectors can foster discussion and steer the evolution of the DPP towards improved interoperability.

The following methodology has been used to derive the requirements: Starting with a review of existing scientific and grey literature on the DPP (which is presented in Section 2), we identified the key goals (Section 3) and most important stakeholders (Section 4) of the DPP. From the literature we also extracted those requirements that were most often mentioned. We furthermore elicited requirements from practitioners from various sectors by means of a questionnaire, which was filled in by 10 manufacturers and 9 end-of-life actors. In addition, we took part in various initiatives, meetings, talks, and discussions with experts on the DPP, and incorporated the experts' opinions into our requirements analysis. We present the results in Section 5. Section 6 summarizes and concludes the paper.

2 Existing Approaches to the Digital Product Passport

Name	Developed by	Sector	Description	Reference
Product Circularity Data Sheet	PositiveImpaKT	all	exchange of standardized, verified circularity data for products	[7]
Resources Passport	Excess Materials Exchange	all	standardized format to enable sharing of resource-related data	[8]
Digital Lifecycle Passport	ReCircE consortium	all	sharing of human and machine readable data along the product lifecycle based on Asset Administration Shell	[9]
Digital Product Passport	GS1	all	machine readable product data based on persistent identity of the product	[10]
Digital Product Passport	iPoint	all	enables manufacturers to record sustainability data about products in a standardized format	[11]
Digital Product Passport	R-Cycle	plastic packaging	storing of production information for recycling based on GS1 standards	[12]
circularity.ID	circular.fashion	textiles	open data standard to enable circularity in the fashion industry	[13]
GBA Battery Passport	Global Battery Alliance	batteries	tracking lifecycle data for batteries of electric vehicles	[14]
Battery Identity Global Passport	UT-Battelle	batteries	tracking lifecycle data for batteries to improve recycling	[15]
Materials Passport (BAMB)	BAMB EU project	buildings	recording data about materials in buildings for recovery and reuse	[16]
Madaster	Madaster Foundation	buildings	recording data about materials in buildings to enable circularity	[17]
International Material Data System (IMDS)	several car manufacturers	automotive	recording data related to materials used in the automotive industry for compliance checking	[18]
Product Lifecycle Management (PLM)	PLM research	all	management of all product related information and processes across the enterprise through its lifecycle	[19]

Tab. 1: Different approaches to the Digital Product Passport.

Several endeavours exist that are concerned with DPPs or similar concepts (see Table 1). Some pursue a cross-sectoral approach [10, 9], while others are conceptualized in a sector-oriented way [14, 17]. Table 1 gives an overview of some of the most prominent approaches that currently exist. All of them are either at the conceptual level [15], in a prototypical stage

[7, 9] or are being tested in first pilots on the market [13]. Technology readiness will likely increase in the coming years when the DPP becomes mandatory in more and more sectors in the EU, starting with battery passports in 2026 [20]. To foster practical DPP uptake, however, several non-technological challenges must be tackled as well, i.e., stakeholders must agree on common standards, new business models around the DPP must be explored and regulatory conditions must be clearly stated so they can be addressed. This paper is intended to contribute in this regard and will hopefully bring the discussion forward.

3 Goals of the Digital Product Passport

Analyzing the existing literature, we extracted the goals associated with DPP concepts. In detail, the DPP should fulfill the following goals:

- ensure that all stakeholders in the product lifecycle can access and share product-related data [9, 10, 7]
- improve collaboration in the value chain [11, 7, 13]
- foster transparency and traceability of products, materials, components [11, 10, 7, 14]
- enable compliance checking and facilitate certification procedures [10, 18]
- enable tracking of critical materials/substances [10]
- improve circularity of products in terms of the R strategies [9, 7, 13, 16, 17]
- allow to benchmark different products [14]
- track progress in the development of sustainable products [14]

As can be seen, most goals were mentioned multiple times; the improvement of the circularity of products being the most frequently mentioned goal, followed by the goal of fostering transparency and traceability of products, materials, and components.

4 Stakeholders

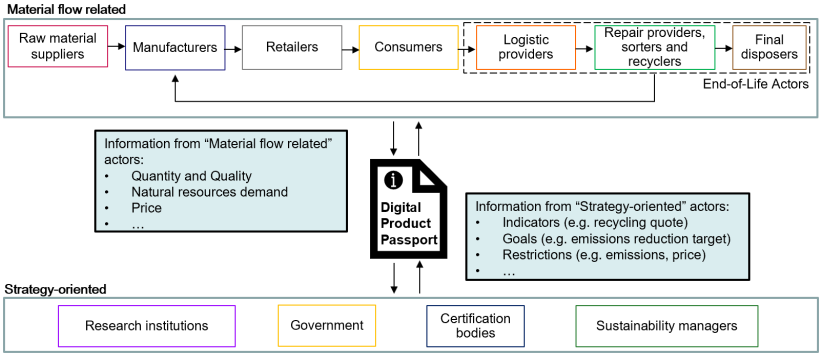


Fig. 1: Potential users of the DPP (based on [21]).

The following section gives an overview of potential users of the DPP. The users are based on and adapted from [21]. They can be divided into two general groups: the material flow related actors and the strategy-oriented actors (see Figure 1). The distinction is based on the *type of interaction* with the DPP, the *type of information* contributed, and the respective *goals of the interaction* with the DPP.

The material flow related actors (e.g., industries, retailers, recyclers) interact physically with the material flow and the necessary processes to handle it. Examples of information provided by the material flow related actors include quantity, quality and composition of a product or material and the energy, water and raw material demand in the production process.

The strategy-oriented actors comprise for example the government, research institutions, certification bodies and sustainability managers. Such actors are able to access material flow information via the DPP for, e.g., better definition of targets, developing strategies, initiating measures and calculating indicators. Such aggregated information is fed back to the DPP, where it serves as a basis for decision makers and for the optimization of processes. This can lead to greater resource efficiency, the reduction of needed resources, or even sufficiency, the elimination of unnecessary processes or reduction of primary resource consumption.

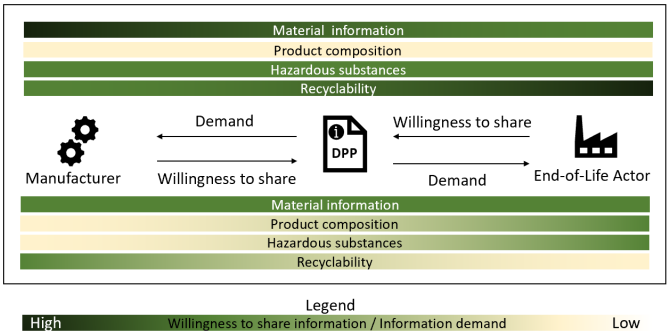


Fig. 2: Results of our questionnaire survey regarding mutual information requirements.

When it comes to mutual information requirements, the results from our questionnaire survey show that demand for information corresponds to willingness to share among manufacturers and end-of-life actors (see Figure 2), with some exceptions: End-of-life actors desire information on product composition and hazardousness, but only some manufacturers are willing to share this information. This may hint at different interests and possible conflicts.

5 Requirements for a Digital Product Passport

The requirements presented in this section were elicited based on our literature review, the results from the questionnaire, talks given by experts and discussions with experts, as mentioned in Section 1.

5.1 Information Requirements

The DPP allows value chain stakeholders to map information and obtain a better understanding of the *composition* of the product, the *environmental impact* of the production and use phase, and the *recycling solutions* at the end of the lifecycle [22]. In most DPP concepts existing so far, the DPP is generated during the production phase of a product by default. However, for recycling-oriented use cases, the DPP can also be generated at a later phase, e.g., during collection or sorting. Depending on the use case, one DPP per individual product might be required (e.g., to record individual manufacturing or usage/repair data). For simple products, one DPP per product model or per batch might be sufficient.

The requirements on a product's DPP depend on a number of criteria, including:

- complexity of the product, i.e., material mixtures, joining techniques, disassembly efforts
- complexity of the lifecycle, i.e., type of use, number of stakeholders during each lifecycle, number of lifecycles
- value of the product or the targeted fractions, i.e., monetary (residual) value or ecological indicators such as embodied energy and material scarcity
- harm potential, i.e., product hazardousness or material toxicity

Manufacturing Data The main responsibility for the subsequent value chain lies in the manufacturing process and its documentation. This should contain the product's composition, the materials and weights used in each component [23]. Additionally, technical data is also significant to be included. This contains details about the manufacturing process (e.g., joining technique, binder), the physical and chemical properties of the materials employed as well as information regarding non-hazardousness or hazardousness to human health or to the environment (e.g., hazard class) [24]. A product's and its components' performance and durability should also be stated. This, together with information on component removeability and replaceability, can help users extend the product's life expectancy and enable optimal product use. The inclusion of the batch number in the DPP is beneficial because it allows to track all products that were manufactured under the same conditions.

Usage Data During the use phase of complex products, all changes to the product should be documented. This includes the documentation of any parts that have been replaced or repaired. The person or organization who makes these adjustments is also responsible for updating the DPP. Frequent product-specific damages and damage to components, especially those that are critical to health and environmental safety, should be recorded to ensure proper handling and enable optimization of the product [25]. Ideally, users will indicate how they intend to dispose of the product when it reaches the end of its life. This, in turn, aids in better planning of the CE and better targeting of waste collection campaigns.

End-of-Life Data The DPP should include documentation on collection, sorting, and treatment during the End-of-Life (EoL) phase. These data, when combined with user input, can help improve waste management. The recording of the achieved collection fraction of the product can be compared with the sales volume. This provides information on how many products are still in use, how many have left the geographical boundary of the sales region or have been otherwise displaced. Furthermore, the applied recycling method should be specified, and the process's output streams should be recorded. This information can also be fed back to the manufacturer to facilitate product (re-)design for improved circularity.

Lifecycle Data Sales volume can be used to anticipate how much waste to expect at any given time and the amount of resources that could be recycled. This facilitates economic and CE planning. However, these are variable figures that depend on a variety of factors such as the life expectancy of a product, user behaviour, and recycling potential. Furthermore, the DPP should include guidelines for storage and usage of the product, which are provided to users in order to minimize potential health or environmental risks. The DPP also serves as a tool for achieving the Sustainable Development Goals (SDGs) [26]. Therefore, information on the environmental impact of a product should be disclosed, which aids consumers with their purchasing decisions [27]. Moreover, a social lifecycle assessment (SLCA) can be incorporated to assess and identify potential social impacts throughout the lifecycle [28].

5.2 Requirements for Successful Collaboration

Collaboration, trust and transparency are the most difficult aspects of the CE. Therefore, all stakeholders in the value chain must work together with openness to achieve long-term success [29]. Data availability and transparency are essential throughout a product's lifecycle to indicate its origin, location, material properties and characteristics as well as end-of-life treatment and recycling, among other things – especially if a global CE is to be realized. In resource planning and sustainable procurement practices, materials traceability is critical. In many circumstances, tracking information is limited, particularly in developing countries with weak governance structures. Digital technologies such as a DPP aim to enhance transparency and traceability [30].

5.3 Identification

A challenge in the pursuit of digitization of product information is the linkage between the physical product and the associated information that resides with various stakeholders. A unique product identification is critical for storing product-related information because it allows the physical products to be connected to their specific information. Additionally, this identification must be recognizable on a global scale [19]. Ideally, the identifier should not only be able to be carried on a data carrier like a QR code but also have a uniform

resource locator (URL) on the internet. Moreover, linking the identifier with other common existing identifiers such as the European Article Number (EAN) or other GS1 standards [10] ensures traceability of the information added to the DPP from production to disposal of the product. In addition, such information should be linkable to certain stakeholders (e.g., manufacturers) to achieve accountability [24]. The storage of the DPP and access to it are further topics that need to be considered, but are not the focus of this paper.

5.4 Incorporation of Legal Obligations

The legal and regulatory framework for the DPP is slowly emerging. At present, the main driver for this is the European Commission. The DPP has been mentioned and conceptualised in various strategies and other documents relating to the twin transition – the European Green Deal and the EU's industrial strategies.

The key document is the Proposal for Ecodesign for Sustainable Products Regulation (ESPR) [31] published in March 2022. In this document and its annexes, the European Commission proposes a definition of the DPP, its scope, and the informational requirements of the DPP, which specifically relate to CE purposes and the declaration of substances of concern (i.e., all substances that may negatively affect reuse or recycling of a product). According to the current proposal, every product containing substances of concern that is being introduced to the EU's market will be obliged to have a DPP. There is, however, little mentioning of technological specifications for the DPP system in the ESPR. The current proposal suggests a central registry which will at least contain a unique identifier per product, item series or batch, respectively, but does not make explicit which other information needs to be contained within a DPP. However, other rules and standards can be used for this, such as DIN SPEC 91406 [32], which provides the normative background for unique identifiers.

While the regulatory framework is still sketchy and at an early stage, there are also some further aspirations for obligatory content and use beyond CE. One of them refers to the DPP as a single point of reference and entry for information that companies have to declare. Hence, instead of having to fill in a number of databases, like SCIP [33], REACH, RoHS [34] or EPREL [35] in the European context, which can be cumbersome and costly, the DPP would gather and provide the required information at one single point of truth. Likewise, the provision of information on social compliance and greenhouse gas emissions could be part of a DPP's obligatory content. Another legal item that could be covered by the DPP is the enablement and support of market surveillance via DPPs – they might be used to counter plagiarism and other illegal practices in product manufacturing, for instance.

6 Conclusion and Outlook

In this paper, we have compiled requirements for the DPP based on identified goals and stakeholders. Our compilation constitutes a first proposal and is by no means complete.

Furthermore, the results from our questionnaire are based on a comparatively small number of participants and deserve further investigation. This paper can thus be seen as a basis for a discussion across sectors and industries as to what a broadly applicable DPP concept should look like. Once such a DPP concept has been realized, various applications and services can be built on top of it that can bring the CE forward. Very promising applications are being developed in the field of Artificial Intelligence. Examples are AI-based lifecycle assessment [6] and neural network based waste sorting informed by data from the DPP as proposed by [9]. Future work includes a requirements analysis of the data handling system for the DPP.

Acknowledgement

This work is funded by the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection, project ReCircE, grant number 03EN2353B.

References

[1] European Parliamentary Research Service (EPRS). *Circular economy*. 2021. URL: https://www.europarl.europa.eu/thinktank/infographics/circulareconomy/public/index.html (visited on 06/15/2022).

[2] The European Commission. *The European Green Deal*. 2019. URL: https://eur-lex.europa.eu/resource.html?uri=cellar:b828d165-1c22-11ea-8c1f-01aa75ed71a1.0002.02/DOC_1&format=PDF (visited on 06/15/2022).

[3] The European Commission. *A new Circular Economy Action Plan For a cleaner and more competitive Europe*. 2020. URL: https://eur-lex.europa.eu/resource.html?uri=cellar:9903b325-6388-11ea-b735-01aa75ed71a1.0017.02/DOC_1&format=PDF (visited on 06/15/2022).

[4] German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection. *Digital Policy Agenda for the Environment*. 2020. URL: https://www.bmu.de/DL2444-1 (visited on 06/15/2022).

[5] Thomas Götz, Thomas Adisorn, and Lena Tholen. *Der digitale Produktpass als Politik-Konzept*. Wuppertal Institut, 2021. URL: https://epub.wupperinst.org/frontdoor/deliver/index/docId/7694/file/WR20.pdf (visited on 06/15/2022).

[6] José Oduque de Jesus, Karla Oliveira-Esquerre, and Diego Lima Medeiros. "Integration of Artificial Intelligence and Life Cycle Assessment Methods". In: *IOP Conference Series: Materials Science and Engineering*. Vol. 1196. 1. 2021.

[7] The Circularity Dataset Initiative. *Product Circularity Data Sheet (PCDS)*. 2022. URL: https://pcds.lu/pcds-system/%5C#pcds (visited on 06/15/2022).

[8] Excess Materials Exchange (EME). *Resources Passport*. 2022. URL: https://www.resourcespassport.com/ (visited on 06/15/2022).

[9] Christiane Plociennik et al. “Towards a Digital Lifecycle Passport for the Circular Economy”. In: *Procedia CIRP* 105 (2022), pp. 122–127. ISSN: 2212-8271. DOI: 10.1016/j.procir.2022.02.021.

[10] GS1 in Europe. *Proposed Architecture and Principles for Digital Product Passports*. The Global Language of Business. 2022. URL: https://www.gs1.eu/news/eu-digital-product-passport-revealed-time-to-act (visited on 06/15/2022).

[11] *Digital Product Passport*. 2022. URL: https://www.ipoint-systems.com/de/loesungen/digital-product-passport (visited on 06/15/2022).

[12] R-Cycle. *Projekt R-Cycle*. 2022. URL: https://www.r-cycle.org (visited on 06/15/2022).

[13] circular.fashion UG. *circularity.ID® Open Data Standard Version 2.0*. 2020. URL: https://circularity.id/static/circularity.ID-Standard-Specification-v2.pdf (visited on 06/15/2022).

[14] Global Battery Alliance (GBA). *Battery Passport*. 2022. URL: https://www.globalbattery.org/battery-passport/ (visited on 06/15/2022).

[15] Yaocai Bai et al. “Energy and environmental aspects in recycling lithium-ion batteries: Concept of Battery Identity Global Passport”. In: *Materials Today* 41 (2020). ISSN: 1369-7021. DOI: 10.1016/j.mattod.2020.09.001.

[16] EPEA Nederland BV and SundaHus i Linköping AB. *Framework for Materials Passports*. Buildings as Material Banks (BAMB) European Union project. 2017. URL: https://www.bamb2020.eu/wp-content/uploads/2018/01/Framework-for-Materials-Passports-for-the-webb.pdf (visited on 06/15/2022).

[17] Madaster Foundation. *Madaster*. 2022. URL: https://madaster.de/ (visited on 06/15/2022).

[18] Voith. *IMDS Leitfaden*. 2022. URL: https://voith.com/corp-de/IMDS-Leitfaden.pdf (visited on 06/15/2022).

[19] Kary Främling, Mark Harrison, and James Brusey. “Globally unique product identifiers–requirements and solutions to product lifecycle management”. In: *IFAC Proceedings Volumes* 39.3 (2006), pp. 855–860.

[20] European Commission. *Proposal for a Regulation of the European Parliament and of the Council concerning batteries and waste batteries*. 2020. URL: https://ec.europa.eu/environment/pdf/waste/batteries/Proposal_for_a_Regulation_on_batteries_and_waste_batteries.pdf (visited on 06/15/2022).

[21] Erik G. Hansen, Florian Lüdeke-Freund, and Klaus Fichter. “Circular business models: a typology based on actor type, circular strategy and service degree”. In: *Product Lifetimes and the Environment (PLATE)*. 2021.

[22] TANNER AG. “Der digitale Produktpass, Treiber für die digitale Dokumentation”. In: *ABZ-Magazin* (2021). URL: https://www.tanner.de/wp-content/uploads/2021/08/ABZ_01_2021_web.pdf (visited on 06/15/2022).

[23] Thomas Adisorn, Lena Tholen, and Thomas Götz. “Towards a Digital Product Passport Fit for Contributing to a Circular Economy”. In: *Energies* 14.8 (2021). URL: https://www.mdpi.com/1996-1073/14/8/2289.

[24] Susanne Guth-Orlowski. *The digital product passport and its technical implementation*. 2021. URL: https://medium.com/@susi.guth/the-digital-product-passport-and-its-technical-implementation-efdd09a4ed75 (visited on 06/15/2022).

[25] Conrad Tucker and Harrison Kim. “Predicting emerging product design trend by mining publicly available customer review data”. In: *DS 68-6: Proceedings of the 18th International Conference on Engineering Design (ICED 11)*. 2011.

[26] United Nations Development Programme (UNDP). *Sustainable Development Goals (SDGs)*. 2015. URL: https://www.undp.org/sustainable-development-goals (visited on 06/15/2022).

[27] Otmar Lell, Viola Muste, and Christian Thorun. *Förderung des nachhaltigen Konsums durch digitale Produktinformationen: Bestandsaufnahme und Handlungsempfehlungen*. 2020. URL: https://www.umweltbundesamt.de/sites/default/files/medien/5750/publikationen/2020_11_17_texte_212_2020_digitalisierung_nachhaltiger_konsum_wirtschaftkonsum.pdf (visited on 06/15/2022).

[28] Andreas Moltesen et al. “Social life cycle assessment: An introduction”. In: *Life Cycle Assessment*. Springer, 2018, pp. 401–422.

[29] Dawei Zhang et al. “Sustainable Circular Business Model for Transparency and Uncertainty Reduction in Supply Chain Management”. In: *J. Theor. Appl. Electron. Commer. Res.* 16 (2021), pp. 959–975. DOI: 10.3390/jtaer16040054.

[30] *OECD workshop on international trade and the circular economy*. 2020. URL: https://www.oecd.org/env/workshop-trade-circular-economy-summary-report.pdf (visited on 06/15/2022).

[31] European Commission. *Proposal for Ecodesign for Sustainable Products Regulation*. 2022. URL: https://environment.ec.europa.eu/publications/proposal-ecodesign-sustainable-products-regulation_en (visited on 06/15/2022).

[32] DIN. *DIN SPEC 91406: Automatic identification of physical objects and information on physical objects in IT systems, particularly IoT systems*. 2019. URL: https://dx.doi.org/10.31030/3114151.

[33] European Chemicals Agency. *SCIP*. 2022. URL: https://echa.europa.eu/de/scip (visited on 06/15/2022).

[34] Eva Hink and Hans-Jochen Lueckefett. “REACH and RoHS compliance in the supply chain”. In: *2012 Electronics Goes Green 2012+*. IEEE. 2012, pp. 1–4.

[35] European Commission. *EPREL – European Product Registry for Energy Labelling*. 2022. URL: https://eprel.ec.europa.eu (visited on 06/15/2022).

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1495

Using Transfer Learning for Quality Improved Forecasting of Temporal Agricultural Processes by Adapting Convolutional Neural Networks

Alexander Münzberg[1], Christian Troost[2], Ansgar Bernardi[3]

Abstract: AI-based decision support can help farmers to reach improved productivity in an environmentally sustainable way. Through transfer learning, an existing Convolutional Neural Network is progressively adapted to provide high quality forecasting results using agricultural time series in the context of different locations, growth and soil types, climate zones, and management variations. The delivered results are validated by appropriate statistical methods and show improved prediction accuracy.

Keywords: Decision Support in Agriculture; AI-based Methods; Transfer Learning; Time Series Analysis; Convolutional Neural Network

1 Introduction

Faced with the challenges of a growing world population, new trends in human food preferences, and the increasing demand for renewable and biologically produced non-food raw materials, modern agriculture faces the necessity to produce food and other organic materials in ever-increasing quantity and quality. Increased agricultural productivity so far has met such demands, but resulted in significant environmental burdens, as the over-use of fertilizers, animal manure, and pesticides leaves unwanted residues and endangers insects and wildlife, soil quality, and groundwater. Smart farming approaches, combining increased digitalization of all aspects of agricultural work, situation- and location-specific fine-grained control of agricultural machinery operation, and AI-based data analysis and decision support, promise individualized and precise optimization of all aspects of resource usage in agriculture. This promises to attain optimal production levels while avoiding surplus usage of expensive and potentially harmful fertilizers, pesticides etc., thus significantly reducing the ecological footprint of agriculture. Predicting the actual plant growth (and thus the amount of fertilizer needed) requires broad and detailed knowledge of manifold

[1] DFKI GmbH, Smarte Daten und Wissensdienste, Trippstadter Str. 122, 67663 Kaiserslautern, Deutschland alexander.muenzberg@dfki.de
[2] Universität Hohenheim, Schloss Hohenheim 1, 70599 Stuttgart, Deutschland christian.troost@uni-hohenheim.de
[3] DFKI GmbH, Smarte Daten und Wissensdienste, Trippstadter Str. 122, 67663 Kaiserslautern, Deutschland ansgar.bernardi@dfki.de

doi:10.18420/inf2022_128

parameters of work processes, soil, weather, and their mutual interactions and influences. Ultimately, valid decision support needs to consider a comprehensive view onto the complex biological system and forecast the biological and economic consequences of possible action alternatives.

Over many years, knowledge about agricultural processes has been accumulated. Scientific work, based on long-term human experience and small-plot experimental evaluations, resulted in a solid body of agricultural knowledge. Codified and programmed into simulation systems, such knowledge can be used to calculate the respective consequences of alternative agricultural decisions, thus giving valuable input for decision support. However, given the nature of the knowledge body, such insights usually disregard the specific details of individual locations or situations of a specific farm or field, and any system adaptation requires high engineering efforts. Recently, machine learning approaches using deep neural networks have shown good results in many application domains when faced with numerous parameters and complex interdependencies. However, training data of sufficient quality and quantity are hard to acquire in agricultural settings! The SimLearn project, funded by the German Federal Ministry of Education and Research (BMBF), investigates whether a successful deep learning-based prediction system can be realized by combining synthetic training data generated by a traditional simulation system with subsequent transfer learning using only a small selection of data from the particular farm under investigation. The resulting network shall support decision making in agriculture both for determining actual work details and for long-term decisions. Within the project we rely on the simulation systems MPMAS and MPMAS_XN developed by University of Hohenheim. These systems allow for the rich simulation of resource consumption, biological output, and economic consequences under various assumptions of work processes pursued, fertilizer & pesticide usage, and weather development. To cover a reasonable breadth of the parameter space, data of agricultural processes and technical capabilities were provided by the Association for Technology and Structures in Agriculture (KTBL). The research pursued within SIMLEARN tries to answer three key questions:

- Is a deep learning approach suited to learn the prediction of plant growth parameters over time (and thus the outcome of agricultural work)? In [Mu22] we showed that a CNN is able to mimic and correctly predict the time series of field plant growth events when trained with data from the given simulation system; for details see below.
- Is it feasible to improve and adapt the trained CNN model to the particularities of a given farm by extending and re-training with limited data sets? This investigation of a Transfer Learning approach to the agricultural domain is the core subject of this paper.
- Can we demonstrate that, in the context of a particular farm, the resulting “learning system” delivers predictions of better quality than the generic simulation system? This experimental evaluation is still in progress.

2 Related Work

The SimLearn consortium already addressed in [Mu22] the question of which approaches can be chosen to support farmers with digital systems in decision making. Here, the problem was addressed that even though classical agent-based simulation systems often provide good predictions that can be used for decision-making, it is difficult to consider all the details of an individual farm. New machine learning approaches have shown excellent potential for learning complex patterns and dependencies (such as Artificial Neural Networks and Deep Learning) but the required large training datasets of real processes performed are not available. The approach of the paper was therefore to train a neural network with the time series data of an agent-based simulation system to replicate the input-output functionality of the agent-based system. Based on the input of initial time series data, future events are to be predicted with the help of a convolutional neural network. This was also successful, with limitations due to differences in plant types, seed rotations, and soil textures. Based on the publication described above, this study deals with the adaptation, extended training and evaluation of the machine learning and data flow processes described in [Mu22] with respect to transfer learning. This means that the existing neural network will be adjusted and tested to be able to predict data values not exclusively corresponding to the training data due to additional and unforeseen events in the input time series. Furthermore, the limitations mentioned above shall be solved by adapted encoding within the data representation of plant types and soil textures. The machine learning algorithm adapted to transfer learning will later be used to predict real processes in agriculture.

A so-called convolutional neural network (CNN), was developed, which is trained by input data of agent-based simulation networks and thus should be able to approximately predict the output time series of these simulation networks. By a convolutional layer of the CNN the large number of input values, which is caused by the large number of state variables of the simulated field parcelles, can be bundled. A special pooling layer is used to reduce the dimension of the data. Then, the data is predicted by regression as in conventional multi-layer perceptrons (MLP) (to be read in more detail in [Mu22]. The input time series data must be preprocessed before training the network. Here, the data is recursively added to the CNN. In the prediction step, an initial number of time series, e.g., the first four days, is used to predict the next (fifth day). In the further step, this predicted day is then included in the input of the training and day six is predicted on the basis of days two to five. Thus, the input time window of the training always increases by one day and includes the respective predicted day. Thus, similar to recurrent neural networks (RNN), a certain history of previously predicted output data is always included in the prediction. This increases the prediction performance of the network [Mu22] [We17].

Figure 1 shows an example of the CNN's testplot prediction. The time series of the variable biomass above ground"is shown, the actual plot and the predicted one.

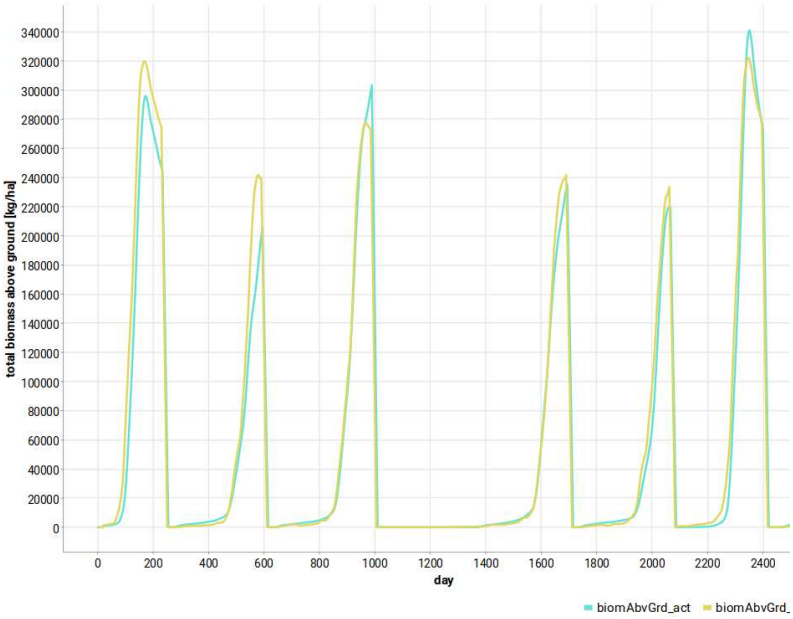


Fig. 1: Actual and CNN-predicted time series biomass above ground"in kg/ha

3 Transfer Learning

Goodfellow, Bengio and Courville describe in [Go16] that transfer learning and domain adaptation refer to the situation in which what is learned in one domain is used to improve generalization in another domain. Furthermore, Torrey and Shavalik in [TS09] describe that transfer learning is the improvement of learning in a new task by transferring knowledge from a related task that has already been learned. The CNN described in chapter 2 is thus successively trained and evaluated from the simulation data of different field plots. In a first step, only plots with the same soil textures and crop sequence are used and gradually further plots with other textures and crop sequence or crop types are added and tested. The cross-validation procedure is used. Here, each dataset (which represents a field plot) from a design point (a pool of datasets) is used step by step once as a test dataset for the prediction. The remaining data sets are used as training data sets in each prediction step (for example, if validation is to be performed using data sets $D1$, $D2$ and $D3$, the data sets $D2$ and $D3$ are first used for training and $D1$ is predicted, then $D2$ is predicted and data sets $D1$ and $D3$ are trained, and so on). Since each of the simulated design points in the SimLearn project contains different compositions of data regarding weather, management and soil or growth conditions, the cross-validation procedure also always trains different values than predicted. Thus, it becomes apparent whether the ML method used is also valid for other compositions of data and states than the trained ones. By repeatedly testing and merging

plots with different ground textures and crop types or sequences, the validity of using the CNN in different areas is examined here.

4 Data Description and Encoding

The MPMAS_XN simulation model is a combination of the mathematical programming-based multiagent system model (MPMAS) and the Expert-N (XN) soil-crop simulator. MPMAS calculates optimal production and investment plans for farms and dynamically simulates their economic outcomes over time. XN simulates crop growth and soil processes over time, including actual yields achieved by the production plans selected by MPMAS [Tr20] [TB16]. As external inputs, MPMAS_XN relies, among others, on a regularly updated KTBL expert data set with technical characteristics of more than 5000 field operations and 2400 machine and equipment types, daily time series of observed or simulated weather data, and multidimensional soil profile data [Mu22].

MPMAS_XN produces the following outputs

- farm-level multivariate data in relational data structures (e.g. optimal production and investment plans for a given farm)
- time series at plot level, i.e. multivariate time series
  - with daily resolution of crop status (biomass, phenology, maturity, nitrogen content, etc.)
  - of soil condition (moisture, carbon stocks, nitrogen, etc.)
  - the adapted management during the season (field work dates, fertilization rates)
  - final yields for a given management plan, initial soil condition, and exogenous weather time series (including temperature, precipitation, humidity, and wind speed).

For the training and validation process carried out in this publication, 15 so-called design points (DP) were used, corresponding to data samples containing time series of 7 field plots of the MPMAS_XN simulation model, each with about 400 of the variables described above and daily time steps for 10 years (per field plot).

The seven plots within a design point have the same soil and weather source (= climate run) and management variation. They differ in the crop sequence. Each DP is assigned 2-3 crop rotations, which are then distributed over the seven plots with their respective different starting positions. In total, one of the crops (Winterrape, Wheat, Winterbarley, Springbarley, Maize) occurs at least once in each DP.

The one-hot-encoding method was used for the labeling of the soil textures as well as for the labeling of the crop species. In advance, the soil textures were assigned to a class according

to sand, silt and clay content, according to the fine soil type diagram in [Du07]. This helps the CNN to distinguish the data for a more accurate prediction [Li19].

5 Results

For each of the 15 design points (DP), the cross-validation strategy was applied to seven time series data sets, each representing a field plot (see Figure 2). After running through the cross-validation of one of the respective DP, a CNN was (re-)trained continuously (across all design points) with the field plots contained within it, which was used for the further cross-validation of the plots. Thus, this CNN gradually adapts to all DP through training. The cross-validations can be used to see to what extent the forecast quality improves as a result of these adjustments.

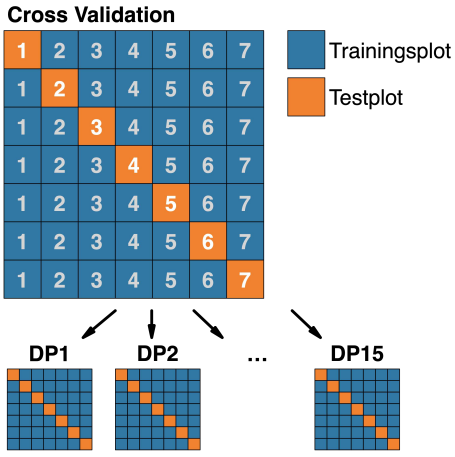


Fig. 2: Cross-validation using 7 field plots across 15 DPs

The root mean squared error function ($RMSE$) shown below was applied as a validation function to the output values (the scaled multivariate CNN output with 10-year time series of the data mentioned in Section 4) of the test plot predictions obtained during cross-validation.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(act_i - pred_i)^2}{n}} \tag{1}$$

Table 1 shows the calculated RMSE values across all field plots and design points (DPs).

Table 1 shows the RMSE values of the test plots predicted during cross-validation, 7 predicted plots per cross-validation step of a design point (DP). As can be seen from the boxplots in Figure 3 and Figure 4, the RMSE median values remain in the same range for

Tab. 1: RMSE values of field plots in design points

	$Plot1$	$Plot2$	$Plot3$	$Plot4$	$Plot5$	$Plot6$	$Plot7$
$DP1$	0.058	0.056	0.060	0.061	0.064	0.058	0.056
$DP2$	0.056	0.048	0.052	0.054	0.050	0.046	0.053
$DP3$	0.051	0.048	0.049	0.144	0.049	0.055	0.056
$DP4$	0.056	0.050	0.064	0.054	0.051	0.048	0.053
$DP5$	0.064	0.056	0.053	0.047	0.050	0.049	0.048
$DP6$	0.055	0.053	0.052	0.127	0.047	0.051	0.046
$DP7$	0.047	0.051	0.048	0.049	0.054	0.049	0.051
$DP8$	0.066	0.050	0.051	0.119	0.048	0.056	0.048
$DP9$	0.065	0.052	0.050	0.045	0.055	0.055	0.057
$DP10$	0.048	0.046	0.047	0.051	0.052	0.052	0.053
$DP11$	0.062	0.057	0.061	0.053	0.051	0.052	0.050
$DP12$	0.052	0.047	0.053	0.043	0.043	0.042	0.042
$DP13$	0.050	0.050	0.051	0.111	0.053	0.052	0.056
$DP14$	0.050	0.050	0.060	0.048	0.047	0.050	0.047
$DP15$	0.052	0.047	0.047	0.057	0.065	0.049	0.054

the continuous CNN adjustment during the cross-validation and still decrease in some cases. It should be noted that plots of DP 3, 6, 8 and 13 contain significant outliers (as can be seen from the yellow marking in table 1 and the box plots in figure 3 and 4). These DPs contain a test plot which is responsible for the outliers and which is not rotated in the crop rotations (not several crops like wheat, barley, maize or rape are included but only the one crop maize is included in the sowing). Overall, it can be said that an already trained CNN will improve its prediction quality by training additional field plots that contain further configurations in terms of climate, soil and management variations. The CNN thus covers a wider range of possible configurations of a field plot with its prediction functionality and delivers at least the same prediction quality as the initially trained network through the extended training in this configuration range.

6 Conclusion

In the present work, a CNN was used to make agricultural forecasts (prediction of time series) based on certain input data. By selecting appropriate crop rotations and management variations, growth and soil conditions can thus be predicted by given soil properties as well as climatic events. It was shown that the CNN used can be adapted with respect to the change of characteristics in terms of transfer learning. By gradually adjusting the CNN, no significant increase in statistical error was detected, or it could be partially reduced. During the validation steps, test plots were predicted whose values were not incorporated into the training in advance on a one-to-one basis. In other words, the CNN was better adapted to specific application areas through training and is now quite capable of making predictions

based on unseen or deviated values. Further experiments will now test whether and to what extent the prediction quality can be improved with data from real farms.

Acknowledgement

The research project SimLearn is funded by the German Federal Ministry of Education and Research under grant FKZ 01|S19073A. Simulations are performed using the high-performance computing resources bwHPC, funded by the Ministry of Science, Research and the Arts Baden-Württemberg, the Universities of the State of Baden-Württemberg, Germany and Deutsche Forschungsgemeinschaft (DFG) through grant INST 35/1134-1 FUGG.

References

[Du07] Duewel, O. et al.: Bodenarten der Böden Deutschlands. FISBo BGR), 2007.

[Go16] Goodfellow, I., et al.: Deep Learning, volume Vol. 1. Cambridge: MIT Press, 2016.

[Li19] Liang, J. et al.: One-hot encoding and convolutional neural network based anomaly detection. Journal of Tsinghua University(Science and Technology), 59(7):523, 2019.

[Mu22] Muenzberg, A., et al.: Machine Learning on Simulated and Real Farm Data Based on an Ontology-Controlled Data Infrastructure. In (Martin, A., et al., ed.): Proceedings of the AAAI 2022 Spring Symposium on Machine Learning and Knowledge Engineering for Hybrid Intelligence (AAAI-MAKE 2022). http://ceur-ws.org/Vol-3121/paper18.pdf, Stanford University, Palo Alto, California, USA, 2022.

[TB16] Troost, C.; Berger, T.: Advances in probabilistic and parallel agent-based simulation: Modelling climate change adaptation in agriculture. In: Sauvage, S., Sánchez Pérez, J.-M., Rizzoli, A. E. (Eds.): Proceedings of the 8th International Congress on Environmental Modelling and Software, July 10-14, Toulouse, France, 2016.

[Tr20] Troost, C. et al.: The Bioeconomic Modelling System MPMAS_XN: Simulating Short and Long-term Feedback Between Climate, Crop growth, Crop Management and Farm Management. In: van Griensven, A., Nossent, J., Ames, D.P. (Eds.) 10th International Congress on Environmental Modelling and Software. Brussels, Belgium, 2020.

[TS09] Torrey, L.; Shavlik, J.: Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques. volume 3. Information Science Reference, New York, USA, chapter Chapter 11, Transfer Learning, 2009.

[We17] Wenpeng, Y. et al.: Comparative Study of CNN and RNN for Natural Language Processing. https://arxiv.org/pdf/1702.01923.pdf, 2017.

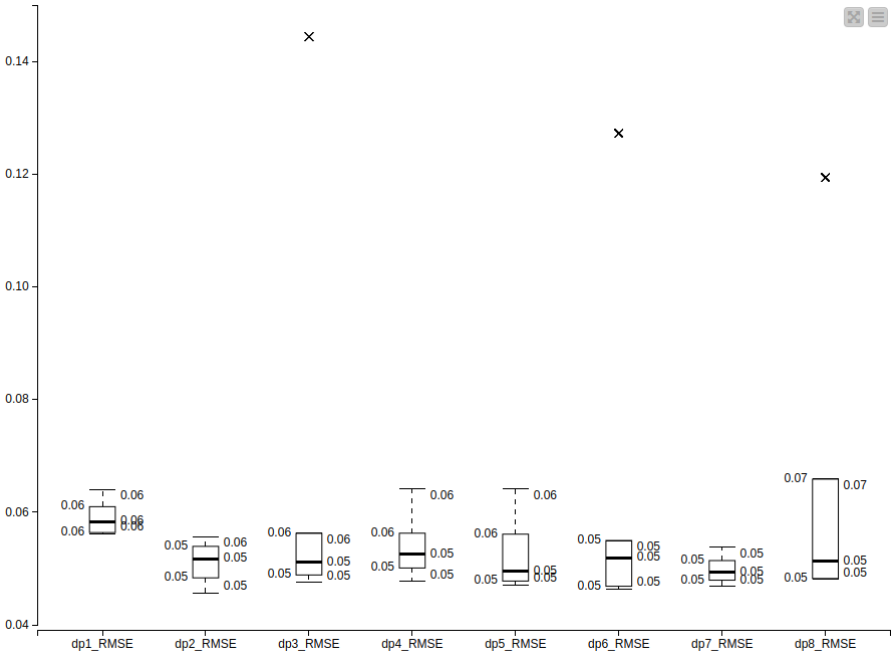


Fig. 3: Box plots with RMSE values of design points 1 to 8

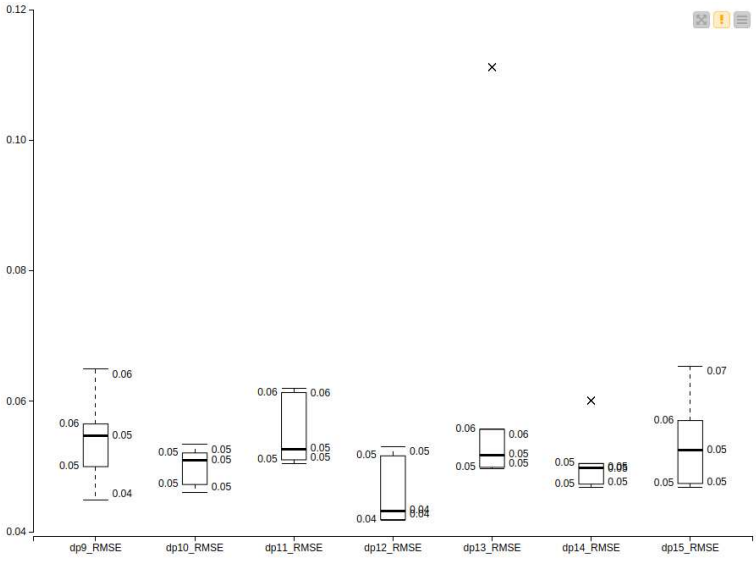


Fig. 4: Box plots with RMSE values of design points 9 to 15

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1505

Comparison of Spatio-Temporal Prediction Approaches of Point-Referenced Environmental Data

Nina Dorffer [1], Julian Bruns [2], Andreas Abecker [3], Stefan Lossow[4]

Abstract: Due to climate change and the effect on human health, there is an urgency to observe and understand the environment. To achieve this goal, knowledge about the development of environmental parameters over time and space is necessary. The analysis of the underlying data can therefore be done with spatial, temporal or even spatio-temporal methods. These methods can also be combined: First spatial,then temporal analyses and vice versa. In this work, we compare the effect of the decision whether to first analyze the data spatially, temporally or both simultaneously. We chose temperature data in Baden-Württemberg, yearly and monthly aggregated, for our comparison.

Keywords: Environmental Computer Science; Spatio-Temporal Statistics; Evaluation

1 Introduction

The environment is everything around us. It is controlled by physical, chemical and other natural forces. Since humans embraced industrialisation different environmental disasters occurred. Two of the most common risks to human health in the developed world are introduced by air pollution and heat stress. Air pollution is defined as an emission of substances in form of aerosols or gas like sulphur dioxide or nitrogen oxides [EA20]. The effect of heat stress is investigated i.e. by [Mu19] in the central and southern parts of Germany. The environment is an important part of human living. Changes in any environmental parameter can cause consequences on human health. Consequently, there is a need to observe our environment.

Environmental data sets often consist of information about the date and location of observation. Based on these information, the analysis of this data can be spatial, temporal or even spatio-temporal. “Spatio-temporal aspects are crucial in order to manage and mine data, to index and retrieve information, and finally to discover and visualize knowledge” [IRT18]. In the last decade, the temporal and spatial resolution of environmental observations increased [GPH16a]. This led to a bigger amount of spatio-temporal datasets. As a result, the issue of spatial, temporal and spatio-temporal analyses of different environmental parameters seems necessary.

[1] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe nina.dorffer@disy.net
[2] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe julian.bruns@disy.net
[3] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe andreas.abecker@disy.net
[4] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe stefan.lossow@disy.net

doi:10.18420/inf2022_129

One restriction of most environmental ground-measured data is that they are mostly measured at fixed points, not over a larger area. In fact, an often appearing request by for example city planners is the availability of data over a whole study area. One way to get such data is from satellites. Usually, the temporal resolution from satellites is currently lower than from ground-stations. A high temporal resolution is needed to calculate a more accurate forecast [SH10]. There are only few possibilities to measure environmental data continously over space and time.

By combining temporal data with a spatial domain it is implied that both variabilities, spatial and temporal, are modelled. As a result, a spatio-temporal analysis is more complicated than just a spatial or temporal analysis. Traditionally the dimensions space and time are treated separately [Li14]. In practice, the location attribute or time attribute is often ignored at first. Consequently, the fields of spatial analysis and time series analysis have developed independently much from on another and the spatio-temporal context is often missing. This leads to the effect that the analysis of space-time data is often done conditionally. To analyse space-time data conditionally means, that first the spatial aspects are analysed then temporal aspects or vice versa. In addition to that, there is often a lack of good software to handle, import, export, display and analyse space-time data which might be the reason for this handling of space-time data. However, new software such as the R package space time and text books are starting to fill this gap ([Bi08], [CW15],[WZMC19], [GPH16a] and [GG10]).

All of this leads to a lack of knowledge regarding the practical application of spatio-temporal analyses - conditional or simultaneously. We aim to compare the impact of the difference in practical applications. By using real world, open datasets from the environmental domain, we select suitable methods and compare their results. We focus on well-known and already implemented methods in the space time and similar R packages, as these would be used by most environmental scientists and practioneers.

2 Related Work

There exists an impressive body of work which compares spatial or temporal interpolations as well as predictions. We have chosen to present a small look into this field.

[Go00] interpolated monthly and yearly aggregated rainfall. Three multivariate approaches using elevation and three univariate methods were compared. The univariate method ordinary kriging got the best result only in the case when the correlation between rainfall and elevation was less than 0.75. [LTP08] compared spatial interpolation methods. They interpolated continuous wind speed surfaces in the study area of England and Wales. The best result was achieved by using co-kriging including the elevation data. Key result is that range, variance and input data values are key parameter, when choosing a method. [Ne14] tested an artificial neural network (ANN) for spatial interpolation. Afterwards, the result was compared with common interpolation methods like inverse distance weighting (IDW)

and ordinary kriging. The result was worse and the computation time is higher than with common interpolation methods.

There are a lot of studies about the comparison of different methods to forecast a time series. For example, [UCK08] compared three ANN methods. The temperature of two research areas in Turkey were investigated. Each ANN produced usable results. Another example is the study by [St12]. They predicted the daily water temperature of the Moisie River, a watercourse in eastern Canada. They concluded that ANN models are promising to forecast water temperature if it was a long time series. The study by [CD14] compared different ARIMA-models to identify a trend of pollutants, temperature and humidity in India. [Te19] compared the forecast by ANN and ARIMA models by calculating a daily forecast for the output energy of wind turbine. In their paper, ARIMA and SARIMA achieved better results than an ANN. Further studies such as [DH06], the M5 competition [MSA22] or textbooks such as [CM09] provide a starting point ofr a broader view into this topic beyond the environmental sciences.

Most studies dealing with spatio-temporal analysis were presenting a solution for a specific problem. They used different approaches, like hierarchical models, ANN or spatio-temporal interpolation. [Mc05] and [SYH09] used a Bayesian hierarchical model to describe and forecast ozone. [An01] used ANN because there was a lack of data in the past. [PI13] calculated a forecast of ozone concentration in the city centre of Athens with ANN. [TdRM20] present an approach for forecasting a large study area. The time series at each station was decomposed. Trend and seasonality were extrapolated. That approach yields a good result for large areas and long-term forecast. [JK17] compared purely spatial methods with isotropic spatio-temporal and anisotropic spatio-temporal approaches. In the example of this study of irradiance data, the anisotropic approach achieved the best results. [Am20] present ä framework for spatio-temporal prediction of climate and environmental data using deep learning"which is evaluated on simulated as well as real-world temperature data.

Most of the studies presented solutions for a specific problem. The comparison of different methods to interpolate or forecast using space and time is not a subject of broader study to the knowledge of the authors. This paper aims to contribute to fill a gap by providing a starting point in comparing and presenting different methods to forecast a study area. We aim to provide first insights when to use which approach, especially for environmental data.

3 Study Area

For our comparison of the different approaches, we use public data for the federal state of Baden-Würrtemberg. Baden-Württemberg is the third largest federal state of Germany by size as well as by population. This state has a variety of different landscapes like the Black Forest, the Rhine plain, the lake of Constance and the Alps in the south (Baden-Württemberg, n.d.). The Swabian Alb in the east determines the landscape.

Data sets for the study area are provided by LUBW (State Office for the Environment, Measurements and Nature Conservation of the Federal State of Baden-Württemberg) as well as by DWD (German national meteorological service). The data used are open and freely available.

The data sets by DWD are combined with the data sets by LUBW to perform a better analysis. The requirements for the data are, that a raster for validation exists and that there are as many stations as possible. Table 1 shows the number of stations that measure meteorological data in Baden-Württemberg divided by LUBW, DWD and both. In addition to that, it is shown whether a raster by DWD from their open data portal, that can be used to validate, is available or not. The temporal resolution for raster providing precipitation is hourly, daily, monthly, yearly, and for years. The temperature raster is available monthly, yearly and for years. For this work, we assume the raster data from the DWD to be correct and not interpolated. While this may introduce an additional bias into the evaluation, we believe that for an early investigation this is acceptable.

	Precipitation	Temperature	Wind velocity	Wind direction	Air pressure
LUBW stations	26	25	26	26	10
DWD stations	131	100	93	93	43
Number of stations	157	125	119	119	53
Raster available	Yes	Yes	No	No	No

Tab. 1: Available meteorological data from LUBW and DWD

Based on the existing data, we use temperature data for our study from the time period 2009 - 2019. We use monthly as well as yearly aggregated data based on the availability. While precipitation data is also feasible, temperature has a field of investigation has a stronger presence in the literature and the results are therefore easier to compare to the literature.

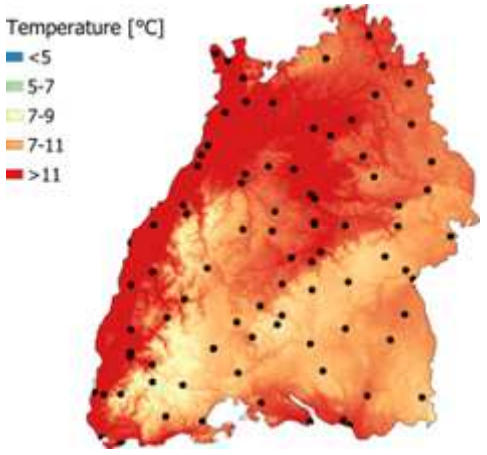


Fig. 1: Temperature raster for the year 2019 and observation stations

4 Methodology and Evaluation

Three different methods are compared in this study. (1) Forecasting using spatial interpolation, then temporal forecasting (S-T), (2) Forecasting using space-time inverse distance weighting (stIDW) and (3) Forecasting using temporal prediction, then spatial interpolation (T-S). For both S-T as well as T-S we compared different approaches for both the spatial and temporal analysis. The best approaches were then chosen and combined in this study. For the temporal models we use an auto-fitted SARIMA model, for the temporal data, we used the auto.kriging function of gstat [GPH16b].

4.1 Forecasting using spatial interpolation, then temporal forecasting (S-T)

In the first step, all previous time steps between 2009 and 2019 are interpolated either yearly or monthly aggregated. The decision, which interpolation method is used, is done by calculating the mean absolute error of the cross-validation. The method resulting in the lowest MAE is chosen. Afterwards, all interpolated raster are laid on top of each other. For each pixel, a time series is created, then for each pixel the value(s) for 2019 is predicted. The workflow can be seen in figure 2. The accuracy of the validation data is calculated for each method. The method with the smallest MAE is chosen to calculate the forecast. To validate the result two different methods are used.

a) In the first validation, the accuracy of the spatial and temporal part is observed individually. We use cross-validation for the spatial anaylsis. The validation of the temporal part is done with a split into training, validation and test data. Therefore the model is trained using data from 2009 – 2016. The validation data used is from 2017 – 2018 and 2019 as evaluation data. This allows us to incorporate a full year with all seasons while still incorporate a large amount of data for training.

b) The result is also validated by calculating the difference to the validation data by DWD. For monthly aggregated data this validation is done for all month. The results are presented for each month and combined by presenting the average.

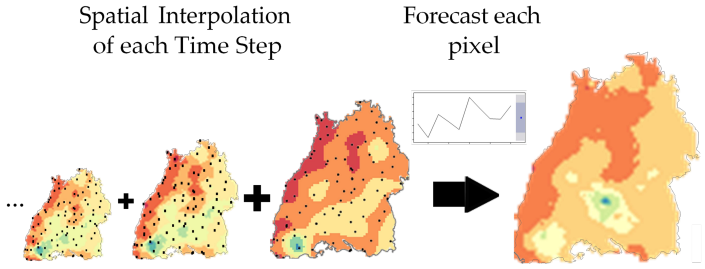


Fig. 2: Workflow to forecast a study area using spatial interpolation, then temporal forecasting

4.2 Forecasting using space-time inverse distance weighting (stIDW)

The method space-time IDW from the package geosptdb is used to calculate the forecast. Space-time IDW is used because the time-intervals were not suitable for space-time kriging in our experiments - during the implementation time, only interpolaton intervalls of less than a day were possible in the package. Based on a preliminary study, the number of neighbours is set to four, the inverse distance power to two and C to one. Therefore the data of the years 2009 till 2018 are used to train the model and to compute a cross-validation. The year 2019 and each month of 2019 was interpolated. To validate the result two different methods are used: a) A leave-one-out cross-validation is calculated and presented. Advantage of this method is, that it is calculated without any information of the year 2019. b) The result is also validated by calculating the difference to the validation raster by DWD.

4.3 Forecasting using temporal prediction, then spatial interpolation (T-S)

The combination is done the other way round in comparison to the previously presented approach for S-T. For each station of the study area values of the wanted time are predicted. The forecast method is chosen by calculating the accuracy of the model for the validation data. The model with the lowest MAE is chosen. Afterwards, the result(s) are interpolated, see figure 3. To decide which method is used to interpolate the MAE of the cross-validation is used. The validation was calculated the same way as for the previous presented method (S-T).

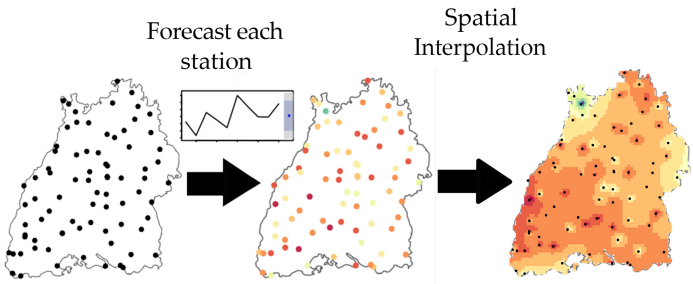


Fig. 3: Workflow to forecast using temporal prediction, then spatial interpolation

5 Results

Firstly, the results for the forecast of the year 2019 are visualized. The test validation leads to the result, that T-S has a low maximal error fig. 4 a. On the other side, the raster validation has a different result. S-T has the lowest MAE and space-time IDW has the lowest maximal error fig. 4 b.

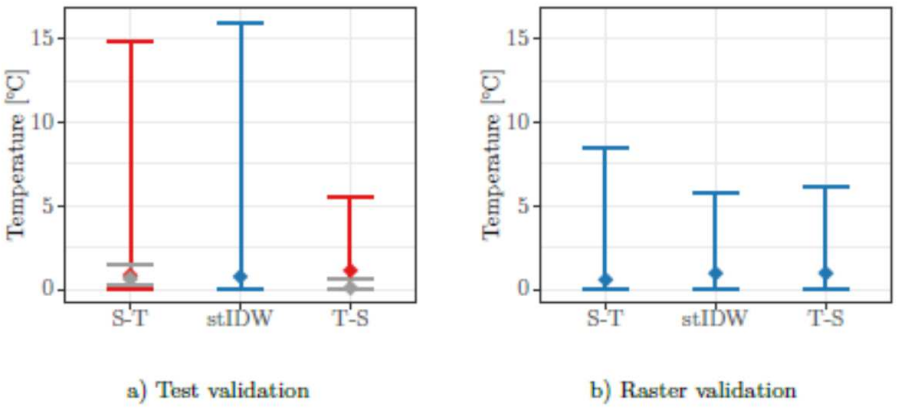


Fig. 4: Comparison of method to forecast of a study area for yearly aggregated temperature. In grey is the temporal error, in red the spatial error and in blue the error of the combined methods.

In fig. 5 the result of the test validation and raster validation of monthly aggregated temperature is presented. In fig. 5a the accuracy of the methods does not differ a lot from each other. It can be said, that the result of the raster validation (fig. 5b) differs from the result of the test validation. In the raster validation, S-T has the lowest maximal error and T-S the lowest mean absolute error. The method space-time IDW has the worst MAE.

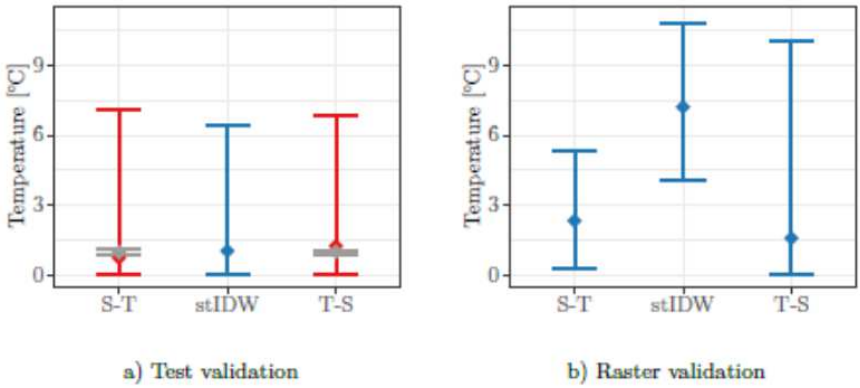


Fig. 5: Comparison of method to forecast of a study area for monthly aggregated temperature. In grey is the temporal error, in red the spatial error and in blue the error of the combined methods.

The results for the raster validation for each month are shown in fig. 6. It can be seen, that the space-time IDW has a MAE about 15 °C in the summer months June, July and August. Reason for that, it that the space-time IDW does not model seasonality. The time part is

modelled with an autoregressive model. The methods S-T and T-S have over the year a MAE lower than 5 °C. However it is noticeable, that the maximal error of T-S can get very high, like in March and April. On the other hand, S-T has the largest maximal error in May with about 6 °C (fig. 6).

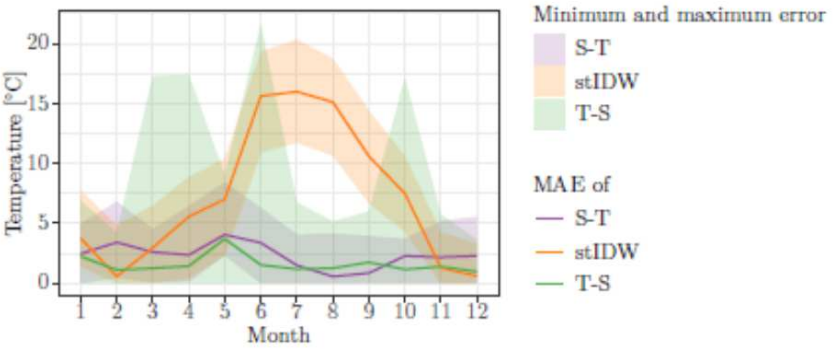


Fig. 6: Raster validation methods to forecast temperature in a study area for monthly aggregated temperature

6 Discussion

Table 2 contains a summary of the results. It is shown which method achieved the lowest MAE, the lowest minimum error and the lowest maximum error. This information is given for the result of the test validation and raster validation. A dash symbolizes that all methods have almost the same result. It can be seen, that results of the test validation and raster validation are not similar. In the test validation the result is often uncertain because all methods led to similar accuracies. Considering the raster validation S-T achieved in most cases the best result. That means the test validation does not help to find the best method to forecast the study area. In fig. 6 it can be seen, that in these cases the reduction methods, S-T and T-S, remains constant and achieves good results in comparison to the extension method space-time IDW.

	Test Validation			Raster Validation		
Data Set	MAE	Max Error	Min error	MAE	Max Error	Min error
Yearly temp	-	T-S	-	S-T	stIDW	-
Monthly temp	S-T	stIDW	-	T-S	S-T	T-S

Tab. 2: Methods with best accuracy in cross-validation and validation to forecast a study area

7 Conclusion

"When data are sparse in space but dense in time, [...] the analysis can be done within the framework of multivariate time series. [...] When data are dense in space and sparse in time, [...] one can work in a multivariate geostatistical setting [...]. In the case, that data are dense in time and space, space-time methods should be used." [Ne17]. While this rule is broadly helpful, it also does not specifiy what the appropiate denseness - the amount of measurements - has to be. In this study we aimed to provide more insight when to use which approach. Three approaches to forecasting a study area were analysed. To interpolate first and make a forecast on each pixel (S-T) led to good forecast results in most investigated cases. It appeared that this approach needed conspicuously long execution time. In comparison, the S-T method needed about 30 times more time than the T-S. And the T-S needed about 30 times more time than space-time IDW based on the work conducted in this paper. Consequently, at this stage, it is either possible to get a result, that is not accurate or to wait until the calculation is done. However, given our dataset, we can say for studies focused on measurement stations - providing dense measurements in time, but not in space - we would advise to first use spatial analysis, followed by the temporal analysis. However, a more in-depth investigation is needed to provide broadly transferable insights.

In this work we used data with a high temporal resolution and a relatively small spatial resolution. This is typical for sensor stations and can also be prevalent in most IoT (Internet of Things) scenarios. We expect this to be a growing use case for spatio-temporal analyses as the IoT use case can be seen in most smart city scenarios - i.e. in the water sector. As sensors are getting cheaper , we will see more application scenarios for those. And while this will lead to an increased spatial coverage, the measurement times will also be reduced to hourly or even measurements every second. But it can also be said, that the method space-time IDW is an improper method for especially monthly aggregated environmental data. That is because, for example, the seasonality is not modelled. The time is modelled using an autoregressive model, which is not sufficient for this use case and therefore most studies in the environmental fields in the next few years.

Another result would have been likely if we investigated the opposite resolutions of the data - low temporal resolution and high spatial resolution. This can be most often seen in the remote sensing field. A follow up investigation to see if the S-T, T-S or spatio-temporal would lead to better results would be of high interest to the authors. Especially of interest is to show or even proof what the conditions are for the better performance of each approach.

For future workother spatio-temporal interpolation methods should be compared, like space-time kriging. This method was not considered in this study because it requires the temporal aggregation not to exceed daily values. In our opinion, not only should more existing methods be compared, but there is also a need for the development of new approaches of spatio-temporal analysis. There do exist a manifold of research paper in regard to natural language processing, even more in regard to image detection (to the benefit of remote sensing), but the use of artifical intelligence for environmental problems with

spatio-temporal data there is still a research gap. Most methods are still provided by statistics - both old and new - but a broader toolset can only benefit the field.

8 Acknowledgement

We would like to thank the reviewers for their intensive as well as constructive feedback. This work was funded by the BMU within the project Nitrat-Monitoring4.0 - Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser (NiMo 4.0), FKZ 67KI2048.

Bibliography

[Am20] Amato, Federico; Guignard, Fabian; Robert, Sylvain; Kanevski, Mikhail: A novel framework for spatio-temporal prediction of environmental data using deep learning. Scientific reports, 10(1):1–11, 2020.

[An01] Antonić, Oleg; Križan, Josip; Marki, Antun; Bukovec, Dragan: Spatio-temporal interpolation of climatic variables over large region of complex terrain using neural networks. Ecological Modelling, 138(1-3):255–263, 2001.

[Bi08] Bivand, Roger S; Pebesma, Edzer J; Gómez-Rubio, Virgilio; Pebesma, Edzer Jan: Applied spatial data analysis with R, volume 747248717. Springer, 2008.

[CD14] Chaudhuri, Sutapa; Dutta, Debashree: Mann–Kendall trend of pollutants, temperature and humidity over an urban station of India with forecast verification using different ARIMA models. Environmental monitoring and assessment, 186(8):4719–4742, 2014.

[CM09] Cowpertwait, Paul SP; Metcalfe, Andrew V: Introductory time series with R. Springer Science & Business Media, 2009.

[CW15] Cressie, Noel; Wikle, Christopher K: Statistics for spatio-temporal data. John Wiley & Sons, 2015.

[DH06] De Gooijer, Jan G.; Hyndman, Rob J.: 25 years of time series forecasting. International Journal of Forecasting, 22(3):443–473, 2006. Twenty five years of forecasting.

[EA20] Emetere, Moses Eterigho; Akinlabi, Esther Titilayo: Introduction to environmental data analysis and modeling, volume 58. Springer Nature, 2020.

[GG10] Gaetan, Carlo; Guyon, Xavier: Spatial statistics and modeling, volume 90. Springer, 2010.

[Go00] Goovaerts, Pierre: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of hydrology, 228(1-2):113–129, 2000.

[GPH16a] Gräler, Benedikt; Pebesma, Edzer J; Heuvelink, Gerard BM: Spatio-temporal interpolation using gstat. R J., 8(1):204, 2016.

[GPH16b] Gräler, Benedikt; Pebesma, Edzer; Heuvelink, Gerard: Spatio-Temporal Interpolation using gstat. The R Journal, 8:204–218, 2016.

[IRT18] Inkpen, Diana; Roche, Mathieu; Teisseire, Maguelonne: , Guest editorial: special issue on Environmental and Geospatial Data Analytics, 2018.

[JK17] Jamaly, Mohammad; Kleissl, Jan: Spatiotemporal interpolation and forecast of irradiance data using Kriging. Solar Energy, 158:407–423, 2017.

[Li14] Li, Lixin; Losser, Travis; Yorke, Charles; Piltner, Reinhard: Fast inverse distance weighting-based spatiotemporal interpolation: a web-based application of interpolating daily fine particulate matter PM2. 5 in the contiguous US using parallel programming and kd tree. International journal of environmental research and public health, 11(9):9101–9141, 2014.

[LTP08] Luo, W; Taylor, MC; Parker, SR: A comparison of spatial interpolation methods to estimate continuous wind speed surfaces using irregularly distributed data from England and Wales. International Journal of Climatology: A Journal of the Royal Meteorological Society, 28(7):947–959, 2008.

[Mc05] McMillan, Nancy; Bortnick, Steven M; Irwin, Mark E; Berliner, L Mark: A hierarchical Bayesian model to estimate and forecast ozone through space and time. Atmospheric Environment, 39(8):1373–1382, 2005.

[MSA22] Makridakis, Spyros; Spiliotis, Evangelos; Assimakopoulos, Vassilios: M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 2022.

[Mu19] Muthers, Stefan; Niemann, Hildegard; Buchholz, Udo; Grabenhenrich, Linus; Matzarakis, Andreas et al.: Schätzung hitzebedingter Todesfälle in Deutschland zwischen 2001 und 2015. Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz, 62(5):571–579, 2019.

[Ne14] Nevtipilova, Veronika; Pastwa, Justyna; Boori, Mukesh Singh; Vozenilek, Vit: Testing artificial neural network (ANN) for spatial interpolation. Journal of Geology & Geophysics, 3(2):01–09, 2014.

[Ne17] Network, RESSTE et al.: Analyzing spatio-temporal data with R: Everything you always wanted to know–but were afraid to ask. Journal de la Société Française de Statistique, 158(3):124–158, 2017.

[PI13] Papaleonidas, Antonios; Iliadis, Lazaros: Neurocomputing techniques to dynamically forecast spatiotemporal air pollution data. Evolving Systems, 4(4):221–233, 2013.

[SH10] Sallis, Philip John; Hernandez, Sergio: Ensemble interpolation methods for spatio-temporal data modelling. In: 2010 Fourth UKSim European Symposium on Computer Modeling and Simulation. IEEE, pp. 132–135, 2010.

[St12] St-Hilaire, André; Ouarda, Taha BMJ; Bargaoui, Zoubeida; Daigle, Anik; Bilodeau, Laurent: Daily river water temperature forecast model with ak-nearest neighbour approach. Hydrological Processes, 26(9):1302–1310, 2012.

[SYH09] Sahu, Sujit K; Yip, Stan; Holland, David M: Improved space–time forecasting of next day ozone concentrations in the eastern US. Atmospheric Environment, 43(3):494–501, 2009.

[TdRM20] Takafuji, Eduardo Henrique de Moraes; da Rocha, Marcelo Monteiro; Manzione, Rodrigo Lilla: Spatiotemporal forecast with local temporal drift applied to weather patterns in Patagonia. SN Applied Sciences, 2(6):1–19, 2020.

[Te19] Tena García, Jorge Luis; Cadenas Calderón, Erasmo; González Ávalos, Gilberto; Rangel Heras, Eduardo; Mbikayi Tshikala, Alain: Forecast of daily output energy of wind turbine using sARIMA and nonlinear autoregressive models. Advances in Mechanical Engineering, 11(2):1687814018813464, 2019.

[UCK08] Ustaoglu, B; Cigizoglu, HK; Karaca, M: Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. Meteorological Applications: A journal of forecasting, practical applications, training techniques and modelling, 15(4):431–445, 2008.

[WZMC19] Wikle, Christopher K; Zammit-Mangion, Andrew; Cressie, Noel: Spatio-temporal Statistics with R. Chapman and Hall/CRC, 2019.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1517

Interaktive Steuerung der Ausführung von KI-Algorithmen in Umweltinformationssystemen über OGC SensorThings

Katharina Emde[1], Matthias Budde[2], Thilo Fischer[3], Tobias Martin[4], Desiree Hilbring[1]

Abstract: Algorithmen der künstlichen Intelligenz (KI) werden über ein weites Spektrum an Domänen zur Entscheidungsunterstützung eingesetzt. Sollen KI-Algorithmen dabei von Fachanwender:innen – die üblicherweise Domänenexpert:innen, oft aber auch IT-Laien sind – wiederkehrend ausgeführt werden, ist ein geeignetes Interaktionskonzept zur Parametrisierung und Steuerung notwendig. Diese Arbeit beschäftigt sich mit der Fragestellung, anhand welcher Kriterien eine dafür geeignete Plattform entwickelt werden muss. Eine Anforderungsanalyse sowie das sich daraus ergebende technische Umsetzungskonzept werden konkret am Beispiel des Umweltinformationssystems „Nitrat-Monitoring 4.0 (NiMo)“ vorgestellt.

Keywords: Umweltinformationssysteme; Künstliche Intelligenz; Entscheidungsunterstützung; Benutzerfreundlichkeit

1 Einleitung

Angesichts großer Herausforderungen bei der Erhaltung der natürlichen Lebensgrundlagen der Menschheit und des sich stetig verschärfenden Klimawandels [Pö22] gewinnen Umweltinformationssysteme (UIS) immer mehr an Bedeutung. Digitalisierung und technischer Fortschritt eröffnen neue Möglichkeiten des digitalen Monitorings unserer Umwelt. Satellitenbilder, aus denen sich beispielsweise der Zustand von Wäldern ablesen lässt [Ya], sowie Luft-, Wasser- und Bodensensoren, helfen uns, ein sich ständig verbesserndes Verständnis komplexer ökologischer Prozesse zu entwickeln. Für die Analyse dieser immer weiter anwachsenden Umweltdatenschätze werden zunehmend Verfahren der künstlichen Intelligenz (KI) entwickelt und eingesetzt. Damit die Ergebnisse solcher Modellentwicklungen nicht ausschließlich einem kleinen Kreis aus Wissenschaftler:innen vorbehalten bleiben und um eine fortlaufende Anwendbarkeit der Algorithmen zu gewährleisten, muss Fachexpert:innen mit geringer technischer Expertise die einfache Parametrisierung und Steuerung solcher Verfahren ermöglicht werden. In diesem Paper werden sowohl die Anforderungen an entsprechende Plattformen als auch ein möglicher Lösungsweg aufgezeigt. Dies geschieht anhand der Plattform, die im Rahmen des Förderprojekts „NiMo“ als Demonstrator entwickelt wird und weiter ausgebaut werden soll.

[1] Fraunhofer IOSB, Fraunhoferstr. 1, 76131 Karlsruhe, Deutschland katharina.emde@iosb.fraunhofer.de
[2] Disy Informationssysteme GmbH, Ludwig-Erhard-Allee 6, 76131 Karlsruhe, Deutschland nimo@disy.net
[3] TZW: DVGW-Technologiezentrum Wasser, Abteilung Wasserversorgung Karlsruher Straße 84, 76139 Karlsruhe
[4] TZW: DVGW-Technologiezentrum Wasser, Abteilung Wasserverteilung, Wasserwerkstr. 2, 01326 Dresden

doi:10.18420/inf2022_130

2 Hintergrund und verwandte Arbeiten

Das Projekt „Nitrat-Monitoring 4.0 - Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser“ (NiMo) [Li21] hat zum Ziel, das Nitratmonitoring im Grundwasser mit Hilfe von Verfahren aus dem Bereich des maschinellen Lernens zu verbessern. Nitrateintrag stellt ein potentielles Problem für die Trinkwassersicherheit in Deutschland dar [LU21], eine Haupteintragsquelle ergibt sich dabei aus der Überdüngung landwirtschaftlich genutzter Flächen. Inhaltlich fokussiert sich das Projekt auf folgende drei Themengebiete:

- Flächenprognosen (Regionalisierungen) von Nitratkonzentrationen im Grundwasser
- Event-Detection-Systeme (EDSe) zur Identfikation von Sonderereignissen (Spitzenwerte, Ausreißer) und Zuständen
- Messnetzoptimierung bestehender Messnetze zur Verbesserung der Abdeckung, der Verteilung und der Beprobungsintervalle

Die Algorithmen des Projekts verwenden eine Vielzahl von Eingangsdaten, neben Zeitreihen zur Nitratkonzentration auch physiko-chemische Parameter (Temperatur, Leitfähigkeit, pH,. . .) und Rasterdaten (Landnutzungsdaten, hydro-geologische Informationen).

Hertweck et al. [He21] analysieren den Einsatz von offenen Standards des Open Geospatial Consortium (OGC) für die Integration eben solcher KI-Algorithmen in UIS. In der Fortführung beschäftigt sich das vorliegende Paper mit der Nutzbarkeit von KI-Algorithmen für Endanwender.

Neben den technischen Aspekten muss auch die Akzeptanz der Nutzer adressiert und hergestellt werden. So argumentiert beispielsweise Kirsch [Ki17], dass KI Systeme innerhalb eines Mensch-Maschine-Interaktionskontextes stets auch von Nutzerseite aus gedacht und getestet werden müssen. Kirsch betont dabei, dass die Qualität von beispielsweise integrierten Erklärungen sich nicht anhand ihrer mathematischen Korrektheit, sondern an ihrer Nützlichkeit für Nutzer:innen messen lassen muss.

Auch wenn es in der Literatur keine vergleichbaren Plattformen gibt, deren Zweck es ist, Fachanwender:innen von UIS durch Einbindung von KI-Algorithmen zu unterstützen, so lässt sich doch auf Untersuchungen zur Nutzungserfahrung beim Umgang mit Verfahren der KI zurückgreifen.

Shin [Sh] und Jacovi et al. [Ja] beschäftigen sich beispielsweise mit Vertrauen in KI-gestützte Systeme. Dabei bemühen sich Jacovi et al. um eine formale Definition von Vertrauen, die erfüllt ist, wenn sich KIs an explizite oder implizite Verträge halten. Die Autoren stützen sich dabei auch auf die Vorgaben der Europäischen Kommission zu vertrauenswürdigen KI-Modellen. Shin fokussiert sich hingegen darauf, wie sich Vertrauen in Systeme aus der Historie der Nutzer:innen ergibt, beispielsweise im Umgang mit Chatbots oder automatisierten Empfehlungen.

3 Methodik

Ausgangspunkt für die nachfolgenden Beschreibungen bilden zwei Anwendungsworkshops, die im Rahmen des NiMo-Projektes stattgefunden haben. Der erste Workshop wurde intern mit den Anwender:innen, die im Projekt beteiligt sind, durchgeführt. Der zweite Workshop war der Öffentlichkeit zugänglich. Der Teilnehmendenkreis bestand dabei aus Fachandwender:innen von Landesbehörden, Trinkwasserversorgungsunternehmen, Ingenieurbüros und praxisnahen Forschungseinrichtungen.

Um Anforderungen an die Plattform zu analysieren, muss zunächst die Gruppe der Stakeholder definiert werden. Diese besteht aus dem Personenkreis, den wir auch für die oben genannten Anwendungsworkshops gewinnen konnten. Also Fachanwender:innen, die notwendiges Domänenwissen (hier: Nitrat) beherrschen. Ihnen soll es ermöglicht werden, ohne tieferes Verständnis der Wirkungsmechanismen von Verfahren des maschinellen Lernens oder künstlicher Intelligenz, deren Ergebnisse als Entscheidungshilfe zu nutzen.

Im Falle von NiMo besteht die potentielle Nutzer:innengruppe beispielsweise aus Mitarbeiter:innen der Umweltämter oder bei Wasserversorgern. Das Interesse liegt dabei hauptsächlich auf Fragestellungen über die historische und aktuelle Situation der Nitratkonzentration: „Wie stellt sich die Verteilung der Nitratwerte im Versorgungsgebiet aktuell dar?“, „Wie haben sich die Nitratwerte entwickelt?“, „Wie sähe die Nitratverteilung aus, wenn sich bestimmte Eingangsparameter ändern sollten?“, „Welche Messstellen werden sinnvollerweise in welcher Frequenz beprobt?“, „Zeichnen sich an bestimmen Messstellen Trends in der Nitratkonzentrationsentwicklung ab?“.

4 Anforderungen aus Nutzer:innensicht

Neben den allgemeingültigen Anforderungen an User Interfaces (UIs), wie Klarheit, Verständlichkeit und einem ansprechenden Layout [Bh] lassen sich für eine Anwendung, die die Zugänglichkeit zu spezifischen KI-Algorithmen für ein Fachpublikum gewährleisten soll, folgende Anforderungen formulieren:

1. **Parametrisierter Start von Algorithmen**: Die wichtigste Anforderung an die Benutzeroberfläche ist die Umsetzung der Kernkompetenz der Plattform, und zwar das parametrisierte Starten von Algorithmen, um neue Prognosen auf Basis von aktualisierten Daten oder neue Szenarien zu berechnen.
2. **Erkennbare Trennung zwischen Daten und Prognosen**: Sobald in einem Datensatz oder einer Grafik die gemessenen Werte sowie durch KI-Algorithmen erzeugte Prognosen gemeinsam auftauchen, muss für die Benutzer:innen eine eindeutige Zuordnung möglich sein.
3. **Einfache Bedienbarkeit**: Die Bedienbarkeit spielt in diesem Kontext eine besonders wichtige Rolle, da Nutzer:innen mit Prozessen arbeiten, deren Funktionsweise sie

nicht im Detail beurteilen müssen. Sie sollen daher durch den Auswahlprozess begleitet werden und in den Möglichkeiten der Nutzereingabe insofern Freiheitsgrade erhalten, dass nur sinnhafte Kombinationen von Parametereingaben möglich sind.

4. **Nachvollziehbarkeit und Reproduzierbarkeit**: Um Vertrauen in die von Algorithmen getroffenen Vorhersagen zu gewinnen, sollten die Ergebnisse (zumindest zu einem gewissen Grad) nachvollziehbar (Gewichtung der Einflussgrößen, die zum Ergebnis geführt haben) und reproduzierbar (identisches Ergebnis bei identischer Nutzereingabe) sein. Die Verfahren sollten außerdem über eindeutige Bezeichnungen identifizierbar und auf detaillierte Beschreibungen der Funktionsweise verweisen, um Transparenz zu gewährleisten.
5. **Hilfe zur Beurteilung der Ergebnisse** Neben den Anweisungen hinsichtlich der Eingabe von Daten, müssen auch Hilfestellungen zur Interpretation und Verlässlichkeit der Ergebnisse bereitgestellt werden, beispielsweise in Form von Hinweistexten, Abweichungen und Konfidenzbändern. Daneben sollten auch Statistiken der Modelltrainings transparent verfügbar sein.

In Abbildung 1 ist eine beispielhafte Oberfläche dargestellt, auf der die erste Parametereingabe für den Messnetzoptimierungsalgorithmus abgefragt wird, integriert in WebGenesis[5].

5 Abgeleitete funktionale Anforderungen

Die in Abschnitt 4 formulierten qualitativen Anforderungen lassen sich in funktionale Anforderungen an die Plattform übersetzen.

1. Für die Algorithmenausführung wird eine Laufzeitumgebung benötigt (qual. Anforderung # 1).
2. Die Plattform muss in ihrer Benutzeroberfläche Eingabemasken für die Parameter der Algorithmen beinhalten (qual. Anforderung #1).
3. Um eine visuell klar erkennbare Abgrenzung zwischen tatsächlichen Messdaten und prognostizierten Werten umzusetzen, muss die Trennung bereits in der Datenbank bzw. dem Datenspeicher vorgenommen worden sein (qual. Anforderung #2).
4. Die überwiegend technische Parameterauswahl für die Algorithmen sollte den Nutzer:innen nicht direkt überlassen werden. Sinnvoller ist es, von den KI-Expert:innen entwickelte Fragenkataloge zu nutzen, die für Fachanwender:innen verständlich sind, und aus deren Antworten sich eine Parametrisierung schlussfolgern lässt (qual. Anforderung #3).

[5] https://www.iosb.fraunhofer.de/de/projekte-produkte/webgenesis-vernetzung-von-daten-zur-entscheidungsunterstuetzung.html

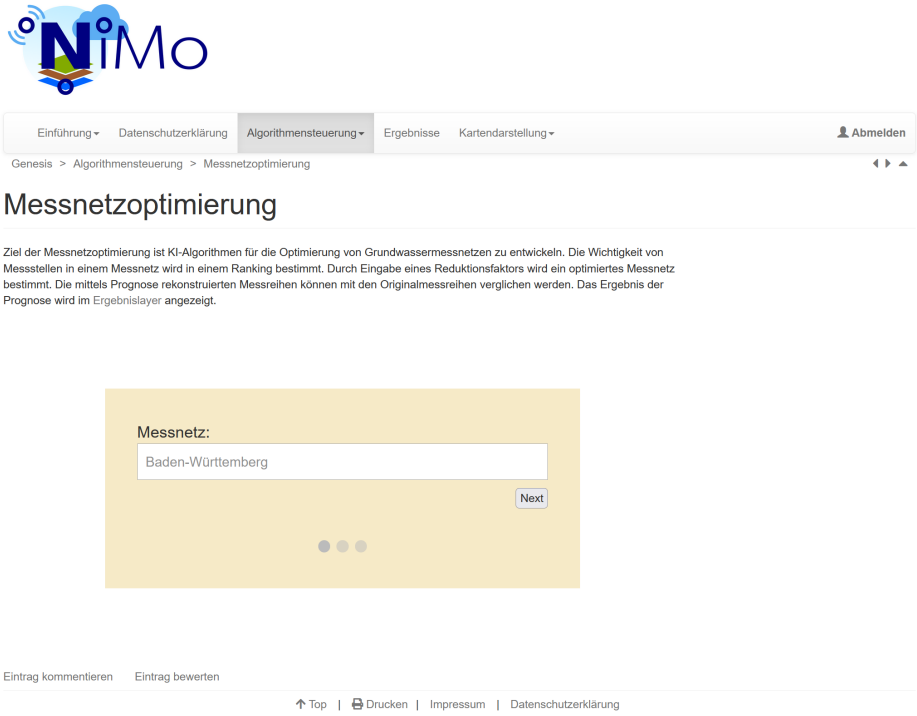


Abb. 1: Parametereingabe für Messnetzoptimierung

5. Ein weiteres Mittel ist die Aufteilung der Plattform in Ansichten, die spezifischen Nutzerrollen entsprechen, um den Kenntnisstand der Anwender:innen bezüglich der eingesetzten KI-Algorithmen zu berücksichtigen (qual. Anforderung #3).
6. Modelle müssen persistiert werden, um jederzeit abrufbar zu sein und dabei die selben Ergebnisse produzieren (qual. Anforderung #4).
7. Die Plattform sollte die Möglichkeit für die Algorithmenentwickler:innen bieten, erklärende Texte sowohl bei der Parametereingabe als auch bei der Ergebnisdarstellung zu ergänzen (qual. Anforderung #5).

6 Exemplarische Technische Umsetzung

Um alle für die Plattform notwendigen Funktionalitäten zu gewährleisten, benötigt das System folgende Komponenten:

- Datenspeicher für Zeitreihendaten (FROST-Server)

- Datenspeicher für Rasterdaten (Geoserver)
- Backend zur Koordination von Containern (PERMA)
- Frontend-Tools für die Interaktion (WebGenesis Instanz, disy Cadenza Instanz)
- Container als Laufzeitumgebung mit Algorithmen (Kubernetes)

Die daraus reslutierende Systemlandschaft ist in Abbildung 2 schematisch dargestellt.

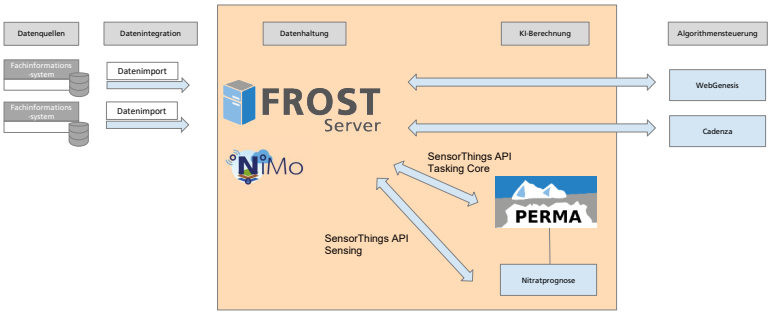


Abb. 2: NiMo Systemarchitektur

Als Speicher für Zeitreihendaten, wie zum Beispiel Nitrat-, Sauerstoff- oder Temperaturwerte, wurde ein FROST®-Server[6] gewählt, für die Rasterdaten ein GeoServer[7]. Mit beiden Tools kann über REST Befehle kommuniziert werden, der Austausch erfolgt über offene Standards der OGC. FROST® im Speziellen ist die Open Source Referenzimplementierung der SensorThings API, die zugeschnitten ist auf die Verwaltung von Sensordaten und damit bestens geeignet für die Speicherung von Grundwassermesswerten. Als Interoperabilitätsstandard für das Internet der Dinge wurde und wird die OGC SensorThings API für eine breite Zahl von Anwendungen eingesetzt, z.B. der Vernetzung städtischer Daten [Fi21] oder der feingranularen Luftqualitätsmessung [Bu17, Li22]. Im Geoserver werden für die Algorithmen benötigte Rasterdaten, wie beispielsweise das Corine Landcover[8] abgespeichert. Diese stehen dort über den ebenfalls durch die OGC definierten, offenen

[6] https://www.iosb.fraunhofer.de/de/projekte-produkte/frostserver.html
[7] https://geoserver.org/
[8] https://www.umweltbundesamt.de/themen/boden-landwirtschaft/flaechensparen-boeden-landschaften-erhalten/corine-land-cover-clcoderhttps://land.copernicus.eu/pan-european/corine-land-cover

Standard Web Map Service (WMS)[9] zur Verfügung. Von dort aus werden die Daten sowohl genutzt, um in den Algorithmen die Analysen zu berechnen, als auch um in der Plattform angezeigt zu werden.

Der Einsatz von offenen Standards ist ein wichtiger Pfeiler bei der Entwicklung eines Systems, das sich — wie die NiMo-Plattform — auf weitere Anwendungsfälle übertragen lassen soll.

Für die Laufzeitumgebung der Algorithmen wurde folgende Lösung gewählt: Die einzelnen Algorithmen bzw. Komponenten der Algorithmen wie Modelltraining und Prognoseberechnung werden mit Hilfe von Docker in einen Container umgewandelt. Die Menge an Containern liegt in einem Kubernetescluster deployed und lässt sich dort über PERMA ansteuern. PERMA bildet einen Aufsatz auf den FROST® Server und macht sich den Tasking Teil der SensorThingsAPI [LK19] zunutze. Dabei werden die Algorithmen selbst als sogenannte virtuelle Aktoren betrachtet, die analog zu realen Aktoren über Tasking Parameter und Tasks verwaltet werden. Dies ermöglicht den Start eines Algorithmus durch das Anlegen eines neuen Tasks per REST Befehl. Da der FROST-Server selbst mit einem Passwortschutz versehen ist, ist auch das Anlegen von Tasks, und damit der Start eines Algorithmus nur für autorisierte Nutzer:innen möglich. Mehr Details zu diesem Lösungsansatz sind bei Hertweck et al. [He21] nachzulesen.

Als Ergänzung zur bereits bestehenden Systemarchitektur kommen mit diesem Artikel die Frontend-Tools WebGenesis[10] und disy Cadenza[11] (Daten-Visualisierung und Analyse) hinzu. Durch sie wird die Interaktion durch Nutzer:innen mit den Daten und Algorithmen ermöglicht. WebGenesis bietet als Entwicklungsplattform für wissensbasierte Systeme die notwendigen Spielräume, um die formulierten Anforderungen umzusetzen. Die Anbindung an die hier verwendeten Standards und vielseitige Darstellungsmöglichkeiten der Daten sind bereits integriert, sowie die Möglichkeit, eigene Inhalte einzupflegen und zu verwalten. Die bereits vorhandenen Tools wurden im Rahmen der NiMo-Plattform genutzt, um Seiten zu konfigurieren, die ausgestattet mit erklärenden Texten die Parametereingabe und Start der Algorithmen ermöglichen, sowie verlinkte Seiten, die die Ergebnisse bereithalten, siehe Abbildung 1.

6.1 Ziele des Systems und Anwendungsfälle

Ziel des Systems ist es, den Nutzer:innen bei der Beantwortung der in Abschnitt 3 beschriebenen Fragestellungen zu assistieren, indem der Zugang zu entscheidungsunterstützenden Systemen gewährleistet wird. Zwei Beispiele für Anwendungsfälle wären ein Mitarbeiter eines Umweltamtes, der die Aufgabe hat, neu erhobene Werte zu überprüfen und als

[9] https://www.ogc.org/standards/wms

[10] https://www.iosb.fraunhofer.de/de/projekte-produkte/webgenesis-vernetzung-von-daten-zur-entscheidungsunterstuetzung.html

[11] https://www.disy.net/de/produkte/cadenza/datenanalyse-software/

Vorselektion und Arbeitserleichterung die Event-Detection der NiMo-Plattform ausführen und die Ergebnisse abrufen möchte oder ein Beschäftigter eines Wasserversorgers, der, um die Beprobungsfrequenz für Messstellen zu bestimmen, die Messnetzoptimierung nutzen möchte.

6.2 Erfüllung der Anforderungen

Das Datenmodell der SensorThings API erlaubt es, jede Entität, neben verpflichtenden Attributen wie „name“ oder „description“, mit flexibleren Attributen, den „properties“, auszustatten. Innerhalb dieses JSON-Objekts kann beliebige Zusatzinformation gespeichert werden, so auch der Ursprung einer Zeitreihe, der entweder in einer der angebundenen Datenquellen oder einem Vorhersagealgorithmus liegen kann. Die Unterscheidung im Datenmodell ermöglicht auch die Unterscheidung in der Darstellung. Ein Beispiel für einen Graphen, der die gemessene sowie eine rekonstruierte Messreihe zu einer bestimmten Messstelle enthält, ist in Abbildung 3 zu sehen.

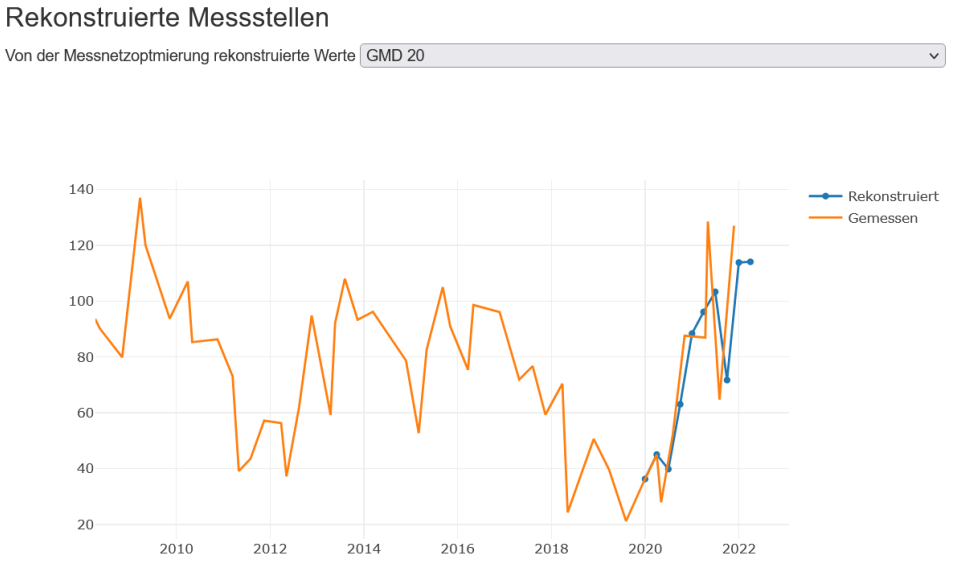


Abb. 3: Graph zweier Zeitreihen: eine stellt die gemessenen Werte dar (orange) und eine die rekonstruierten (blau).

Durch das Zusammenspiel von FROST®, PERMA und dem Kubernetescluster können Algorithmenkomponenten per REST Befehl gestartet werden. Die in Abschnitt 5 geforderte Laufzeitumgebung für die Algorithmen ist über das Kubernetescluster gewährleistet. Um bereits trainierte Modelle für Prognosen wiederholt verwenden zu können, werden diese

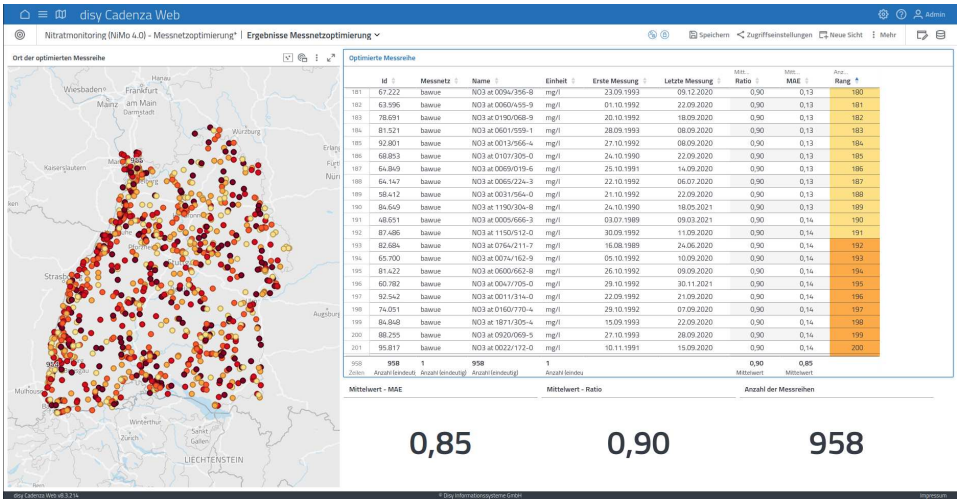


Abb. 4: Ergebnisse einer Messnetzoptimierung in der Location Intelligence Plattform disy Cadenza.

auf Min.io[12], einem Cloudstorage System, persistiert und können von dort abgerufen werden. Die Wiederholungen können entweder ausgeführt werden, um bereits erhaltene Ergebnisse zu reproduzieren, oder um eine Vergleichbarkeit zu schaffen, wenn dasselbe Modell mit unterschiedlichen Daten (z.B. verschiedene Gebiete, oder unterschiedliche Auflösungen) ausgeführt wird. Falls eine Architektur gewählt wird, in der nicht nur aktuelle, sondern auch Vergangenheitsbeobachtungen wieder hinzugefügt werden können, muss für die Reproduzierbarkeit darauf geachtet werden, dass zwischen den verschiedenen Modellläufen keine neuen Daten im Auswertungszeitraum hinzugekommen sind, die das Ergebnis verändern könnten.

Auch WebGenesis als Basis für die Plattform trägt zur Erfüllung einiger Anforderungen bei. Die Möglichkeit, eigenen Inhalt inklusive eigenem Javascript Code einzupflegen, wurde für die NiMo-Plattform genutzt, um Formulare für die Parametereingabe zu schaffen, die in individuell gestaltbare, erklärende Texte eingebettet werden können. Von hier aus werden die vordefinierten Parameter abgefragt und durch das Anlegen eines parametrisierten Tasks per REST-Befehl in den Container und das ausführende Programm darin weitergegeben. Die bereits vorhandene Implementierung der Nutzerverwaltung wird in der NiMo-Plattform genutzt, um die Anforderung umzusetzen, Nutzer:innen mit unterschiedlichem Kenntnisstand unterschiedliche Freiheitsgrade beim Start der Algorithmen zu gewähren. Eine weitere Anwendungsmöglichkeit von WebGenesis besteht darin, Ontologien zu hinterlegen und so eine Wissensbasis aufzubauen. Diese kann durch ihre Verknüpfungen dazu verwendet werden, die Parameterauswahl zu leiten. Im aktuellen NiMo-Projekt befindet sich dieser Teil der Anwendung noch in der Umsetzungsphase. Die Ablage aller Daten, sowohl Inputdaten als

[12] https://min.io/

auch die von Algorithmen erzeugten Prognosen, führt zu der Möglichkeit, nahezu beliebige Graphiken aus dem Datenvorrat zu erstellen und dabei auch Inputgrößen im Vergleich zu Ergebnissen darzustellen und so verschiedene Zusammenhänge sichtbar zu machen. Dies kann in der hier konzipierten Plattform auch durch die Nutzer:innen selbst geschehen, Funktionalität hierfür ist in WebGenesis und disy Cadenza auch bereits vorhanden.

7 Zusammenfassung und Ausblick

Die beschriebene Konzeption einer Plattform, die es Fachanwender:innen ermöglicht, die Vorzüge von KI zu nutzen, ist zugeschnitten auf den speziellen Fall des Projekts NiMo. Dennoch lassen sich die meisten Fragestellungen auf andere UIS übertragen, denn auch eine Anwendung, die KI-Algorithmen zu beispielsweise Wetter- oder Luftdaten enthält, muss den in Abschnitt 4 formulierten Anforderungen standhalten. Erkennbare Trennung zwischen Daten und Prognosen, einfache Bedienbarkeit, Nachvollziehbarkeit und Hilfestellungen zur Beurteilung der Ergebnisse sind Anforderungen, die domänenübergreifend gelten und sich nicht auf hydrogeologische Anwendungen alleine beziehen. Um die Übertragbarkeit nicht nur in den Anforderungen sondern auch bei der technischen Umsetzung zu gewährleisten, setzt die NiMo-Plattform — genau wie die NiMo-Gesamtarchitektur, auf den Einsatz von offenen und weitverbreiteten Standards, wie beispielsweise die des OGC. Im weiteren Verlauf des Projekt werden die hier entwickelten und vorgestellten Konzepte weiter implementiert und zu einem Demonstrator ausgebaut, der zu Projektende bei einem Abschlussworkshop mit Mitgliedern der Zielgruppe präsentiert und getestet wird.

Acknowledgement

Diese Arbeit wurde im Rahmen des vom Deutschen Bundesministerium für Umwelt, Naturschutz, Nukleare Sicherheit und Verbraucherschutz (BMUV) geförderten Projekts „NiMo 4.0“ [Li21] (Förderkennzeichen 67KI2048) erstellt.

Literaturverzeichnis

[Bh] Bhaskar, N. U.; Naidu, P. P.; Babu, S. R. C.; Govindarajulu, P.: General Principles of User Interface Design and Websites. International Journal of Software Engineering (IJSE), 2.

[Bu17] Budde, Matthias; Riedel, Till; Beigl, Michael; Schäfer, Klaus; Emeis, Stefan; Cyrys, Josef; Schnelle-Kreis, Jürgen; Philipp, Andreas; Ziegler, Volker; Grimm, Hans et al.: SmartAQnet: remote and in-situ sensing of urban air quality. In: Remote Sensing of Clouds and the Atmosphere XXII. Jgg. 10424. SPIE, S. 19–26, 2017.

[Fi21] Fischer, Michael; Gras, Pierre; Löwa, Sonja; Schuhart, Stefan: Urban Data Platform Hamburg: Integration von Echtzeit IoT-Daten mittels SensorThings API. ZfV-Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv 1/2021), 2021.

[He21] Hertweck, Philipp; van der Schaaf, Hylke; Hilbring, Desiree; Weis, Jonas; Liesch, Tanja; Budde, Matthias: Integration von KI-Algorithmen in Umweltinformationssysteme mittels SensorThings API. In: INFORMATIK 2021. Gesellschaft für Informatik, Bonn, S. 301–312, 2021.

[Ja] Jacovi, Alon; Marasović, Ana; Miller, Tim; Goldberg, Yoav: Formalizing Trust in Artificial Intelligence. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. ACM.

[Ki17] Kirsch, Alexandra: Explain to whom? Putting the User in the Center of Explainable AI. In: Proceedings of the First International Workshop on Comprehensibility and Explanation in AI and ML 2017 co-located with 16th International Conference of the Italian Association for Artificial Intelligence (AI*IA 2017). Bari, Italy, 2017.

[Li21] Liesch, Tanja; Bruns, Julian; Abecker, Andreas; Hilbring, Désirée; Karimanzira, Divas; Martin, Tobias; Wagner, Martin; Wunsch, Andreas; Fischer, Thilo: Nitrat-Monitoring 4.0 – Intelligente Systeme zur nachhaltigen Reduzierung von Nitrat im Grundwasser. In (Reussner, Ralf H.; Koziolek, Anne; Heinrich, Robert, Hrsg.): INFORMATIK 2020. Gesellschaft für Informatik, Bonn, S. 1069–1079, 2021.

[Li22] Li, Chaofan, et al.: SmartAQnet 2020: A New Open Urban Air Quality Dataset from Heterogeneous PM Sensors. ProScience 8, 2022.

[LK19] Liang, Steve; Khalafbeigi, Tania: OGC SensorThings API Part 2–Tasking Core, Version 1.0. Open Geospatial Consortium, 2019.

[LU21] LUBW Landesanstalt für Umwelt Baden-Württemberg Referat 42 - Grundwasser: , Grundwasser-Überwachungsprogramm Ergebnisse 2020. https://pudi.lubw.de/detailseite/-/publication/10353-Grundwasser-%C3%9Cberwachungsprogramm._Ergebnisse_2020.pdf, 2021.

[Pö22] Pörtner, H.-O.; Roberts, D.C.; Tignor, M.; Poloczanska, E.S.; Mintenbeck, K.; Alegría, A.; Craig, M.; Langsdorf, S.; Löschke, S.; Möller, V.; Okem, A.; (eds.), B. Rama: , Climate Change 2022: Impacts, Adaptation, and Vulnerability, 2022.

[Sh] Shin, Donghee: How do people judge the credibility of algorithmic sources? AI & SOCIETY, 37(1):81–96.

[Ya] Ya'acob, Norsuzila; Azize, Aziean Binti Mohd; Mahmon, Nur Anis; Yusof, Azita Laily; Azmi, Nor Farhana; Mustafa, Norfazira: Temporal Forest Change Detection and Forest Health Assessment using Remote Sensing. 19:012017.

7th GI/ACM I4.0 Workshop on Industrial Automation and Control Systems

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1531

Modeling Security Controls and System Assets As Autonomous Planning Tasks

Mahmoud Khalaf[1], Ludger Peters[2], Karl Waedt[3]

Abstract:

Safety I&C (Instrumentation & Control) and Operational I&C programmable digital systems are growing in complexity at a rapid pace while system designers, project architects, and cyber-security engineers work tirelessly to ensure the safety of the systems by complying with long lists of rules and regulations dictated by relevant regional & international standards [20; BQB17]. These standards are updated, withdrawn, replaced by a revised edition, and amended fairly frequently. Guaranteeing new and existing I&C system assets are still compliant is arduous, expensive, and time-consuming. In this paper, we propose an approach that assists in security modeling and system design by formulating the security controls and I&C assets in PDDL domain (Planning Domain Definition Language by D. McDermott et al. [D 98]). A domain-independent general purpose planner can explore the *state space* and provide a *deterministic plan* that transforms the *initial state* into the *goal state*. The *initial state* in this context can be the security threats from which the system needs protection. The *goal state* can be reaching a specific security degree (S1, S2, S3), satisfying risk management requirements, availability requirements, performance requirements, or a combination of them Tellabi et al. [Te18].

Keywords: Security Controls; I&C System Assets; AI Planning; PDDL; First-order Logic

1 Motivation

> "When you are writing a document for a human being to understand, a human being will look at it and nod his head: 'Yeah, this makes sense'. But then there are all kinds of ambiguities and vagueness that you do not realize until you try to put it into a computer." — Donald Knuth

2 Background

PDDL has gone a long way in recent years and is being incorporated in numerous fields (e. g. penetration testing, security assessment, robot mapping, urban planning, and traffic control). Mainly due to it being the lingua franca for the International Planning Competition[4], and as a result, there is a rich eco-system of tools and services that help the modelers and practitioners to define tasks in a declarative manner intuitively.

[1] Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany khalaf@cs.uni-freiburg.de
[2] Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany lpeters@informatik.uni-freiburg.de
[3] Framatome GmbH, Erlangen, Germany Karl.Waedt@framatome.com
[4] https://www.icaps-conference.org/competitions/

doi:10.18420/inf2022_131

3 Planning Semantics

Here we will primarily focus on using *Fast Downward* Helmert [He06]. *Fast Downward* is a domain-independent classical planning system and considered to be the state-of-the-art planner, usable both for research and practical applications Nebel; Mattmüller [NM19].

Fast Downward can be substituted for another PDDL planner with minimal-to-no modification to the previously-defined domain. This might be desirable to improve runtime performance or describe the premise in a more expressive syntax.

3.1 Formal Definition

Every task here is being tackled from the perspective of classical planning. Deterministic dynamics, full observability, and finite horizon. A deterministic planning task in finite-domain representation is a 4-tuple: $\Pi = \langle V, I, O, \gamma \rangle$ where V is a finite set of finite-domain state variables, I is an initial state over V, O is a finite set of finite-domain operators over V, and γ is a formula over V describing the goal states[5].

3.2 Planner Input

The `problem.pddl` file can be machine-generated. Such that objects (V) can be populated from an external source[6]. The initial state (I) of the problem statement is generated based on an aggregation of sensor data from the target environment, which maps to a list of initially true state variables. The operator can manually specify the goal state (γ).

The `domain.pddl` file for a specific standard is a one-time investment that requires efficient knowledge representation on both the security field and PDDL modeling. Ideally, cybersecurity experts and PDDL modelers work together to encode a standard into the target domain, possibly with the aid of some custom high-level constructs[7].

Each security degree should have its own `domain.pddl` file. the differences between them are mostly in the precondition of the *actions*. Lower security degrees usually have fewer preconditions. A modeler will start with the highest security degree and comment out the precondition in a new `domain.pddl` file according to each security requirement.

3.3 Plan Validation

A validator (`VAL` by Howey et al. [HLF04]) can be used to make sure that previously generated plans are still valid in case of an update to the `domain.pddl` file. For every new

[5] Multiple plans that reach the same *goal state* may exist.
[6] e. g. database or file
[7] e. g. state-space visualization, action usability detection

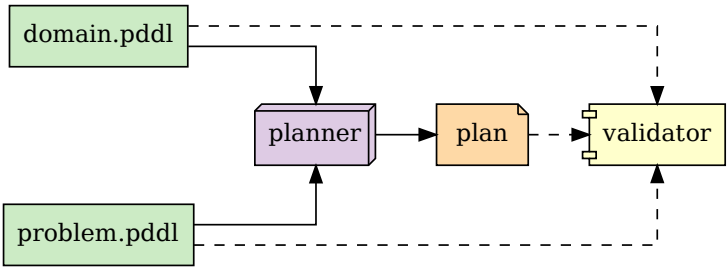


Abb. 1: Plan synthesis overview

amendment in the standard. Accordingly, the `domain.pddl` file needs to updated. The next step will be to check if the plan used to reflect the status quo is still compliant with the amends. The validator succeeds if it can reach the goal state from the initial state using the steps in the plan given the updated domain. If the validator fails, then a new plan needs to be generated from the planner and adjusted in the environment to reflect the new plan accordingly.

4 PDDL Encoding

Before starting any planning task; types, predicates and *actions* have to be defined for the planner in the `domain.pddl` file. Types should follow a reasonable hierarchy to provide logical separation and improve grounding performance. Predicates describe the current premise by accepting zero or more typed parameters. These predicates represent the attributes and relations between their arguments. Any undeclared predicate is considered false[8]. *Actions* are the possible transitions the planner considers to reach the goal state.

4.1 Type Hierarchy

Each I&C system asset in the environment will correspond to an object in the `problem.pddl` file. For example, A computer is a `DevelopmentResource`; Therefore, any action or predicate that accepts a `DevelopmentResource` as one or more parameters; will also accept any descendants of a `DevelopmentResource` in the type hierarchy.

[8] Closed world assumption

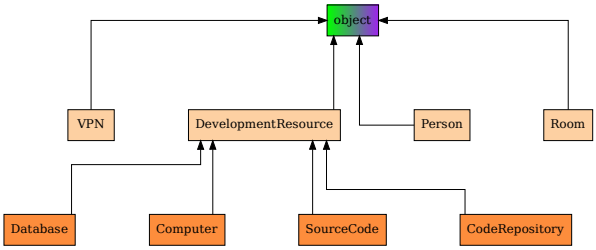


Abb. 2: Type Hierarchy
`domain.pddl`

```
(:objects
    alice bob       - Person
    mysql1          - Database
    thinkpad440     - Computer
    open_sw         - SourceCode
    gitlab          - CodeRepository
    internalIKED    - VPN
)
```

List. 1: objects
`problem.pddl`

4.2 Predicate Definition

A modeler can describe the current premise implicitly or explicitly. `allow_share_resource alice codev1` and `forbid_share_resource alice codev1` are used to tell the planner that proper countermeasures need to be taken to prevent sharing of confidential resources from an authorized party to an unauthorized party. Failing to define the `allow*` or `forbid*` predicates in the `problem.pddl` implies that there is a security risk of sharing confidential resources with unauthorized parties. Moreover, the modeler needs to allow or forbid it explicitly. Otherwise, the modeler might have accidentally overlooked a potential security risk, and the planner will act as a reminder.

In other cases, the `forbid` can be implicit. for example `allow_room_access bob controlroom`. all personnel defined in the `problem.pddl` file don't have access to the control room except bob, and there is no need to explicitly `forbid` each and every individual.

4.3 Transitive Relations

It is a matter of time before running into a situation where the modeler might prefer encoding a transitive relation implicitly. For example, if you access a version control system, you implicitly access the source code on it. Such relation can be represented as a derived predicate (Edelkamp; Hoffmann [EH04]):

```
(:derived
  (derived_authorized_resource_access ?p - Person ?c - DevelopmentResource)
  (and
      (exists
          (?d - DevelopmentResource)
          (and
              (stored_in ?c ?d)
              (or
                  (authorized_resource_access ?p ?d)
                  (derived_authorized_resource_access ?p ?d)))))
)
```

List. 2: Derived Authorized Resource Access

4.4 Action Schema

Actions consist of typed parameters, preconditions, and effects. When the current state satisfies the precondition of a particular action, the planner may consider this action and execute a *state transition* by mutating the predicates in effect. This process keeps repeating until reaching the goal state or a dead end[9].

4.4.1 Blue Team Perspective

Blue teams aim to maintain the security of assets and comply with the relevant standards. An *Action* in this context can be implementing a specific security protocol, upgrading an existing asset, or eliminating a particular threat. Failure to reach the *goal state* might indicate the absence of critical security elements that need to be defined from the security catalog.

The plan in List. 4 guides the blue team to follow the proper protocol for installing or upgrading a software inside an environment operating under strict security constraints:

```
(define (problem install_upgrade_sw)
    (:domain IAEA_MRT_330)
    (:objects
        sw                          - Software
        engineer                    - Engineer
        technician                  - Technician
        licensedOperationsStaff     - LicensedOperationsStaff
        computerSecuritySpecialist  - ComputerSecuritySpecialist
        flashdrive                  - RemovableMedia
        removableMediaIssuer        - RemovableMediaIssuer
    )
    (:init
        (not (is_deliver_sw sw))
        (has_cabient_key licensedOperationsStaff)
    )
    (:goal
        (and (installed_sw sw)) )
)
```

List. 3: Install Software Protocol (`problem.pddl`)

```
0 : (deliver_sw sw engineer)
1 : (verify_hash sw technician)
2 : (test_plan sw computerSecuritySpecialist)
3 : (regression_test sw engineer)
4 : (request_media sw flashdrive engineer)
5 : (issue_media flashdrive engineer removableMediaIssuer)
6 : (transfer_sw sw technician flashdrive)
7 : (scan_sw sw technician flashdrive)
8 : (cabinet_key sw engineer licensedOperationsStaff)
9 : (unblock_cabinet engineer)
10 : (install_sw sw engineer)
11 : (functional_test sw engineer)
12 : (lock_cabinet sw engineer)
13 : (return_keys engineer licensedOperationsStaff)
14 : (scan_media flashdrive engineer)
15 : (return_media flashdrive engineer)
16 : (sanitize_media flashdrive removableMediaIssuer)
17 : (done_install_sw sw flashdrive)
```

List. 4: Install Software Protocol (plan)

4.4.2 Red Team Perspective

Red teams aim to evaluate the security of a system by identifying weaknesses and potential security threats and providing feedback to the blue team to mitigate any security risk. An *action* in this context can be exploiting a misconfigured asset, disclosing confidential resources, or elevating access to an unprivileged user. Failure to reach the *goal state* indicates that the planner cannot find[10] a plan that exploits the system's security.

[9] unsolvable task
[10] under the given assumptions

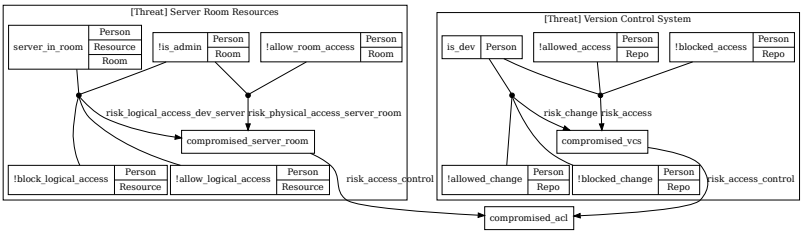


Abb. 3: Compromising Access Control (sample)

```
(:action risk__logical_access_dev_server
    :parameters (?p - Person ?s - DevelopmentServer)
    :precondition (and
        (exists (?r - Room)
            (and
                (server_in_room ?s ?r)
                (not (is_administrator ?p ?r))
                (not (block_logical_access ?p ?s))
                (not (allow_logical_access ?p ?s))
            )))
    :effect (and (compromised_server_room)))
```

List. 5: Logical Access Compromise Action

```
(:action risk__physical_access_server_room
    :parameters (?p - Person ?r - Room)
    :precondition (and

        (not (is_administrator ?p ?r))
        (not (allow_room_access ?p ?r))
        (has_room_key ?p ?r)

    )

    :effect (and (compromised_server_room)))
```

List. 6: Physical Access Compromise Action

Abb. 3 shows some possible ways a malicious actor can compromise a server room according to the standard [20]. Each node corresponds to a predicate that needs to be satisfied. Inside each node are the parameter types that the predicate accepts. Each edge is an action the planner might consider to set the next predicate to true. `compromised_acl` is a predicate that take no parameters and can be reached when either by `compromised_server_room` or `compromised_vcs` is true.

List. 5 and List. 6 are the PDDL representation of the possible ways an attacker can compromise the access control policy according to the standard[20].

In List. 8 the planner managed to detect a breach in the protocol and is warning the user that an explicit confirmation that `alice` cannot share or transfer data to `bob` is needed:

> Authorized personnel must not share or transfer the source code or application software with unauthorized individuals. — *IEC 63096; Nuclear Power Plants – Instrumentation, Control and Electrical Power Systems Security Controls* [20]

```
(define (problem attack_acl)
  (:domain IEC_63096_ACCESSCONTROL_9112_S1)
  (:objects
      alice bob - Person
      mysql1 - Database
      thinkedge450 - Computer
      open_sw - SourceCode
      gitlab - SourceCodeRepository
      internalIKED - VPN
  )
  (:init
      (has_2FA internalIKED)
      (authorized_resource_access alice mysql1)
      (authorized_resource_access bob mysql1)
      (forbidden_resource_access alice thinkedge450)
      (authorized_resource_access bob thinkedge450)
      (authorized_resource thinkedge450)
      (stored_in gitlab thinkedge450)
      (stored_in open_sw gitlab)
      (stored_in mysql1 thinkedge450)
  )
  (:goal
      (and (compromised_access_control)))
)
```

List. 7: Access Control Analysis (`problem.pddl`)

```
0 : (risk__resource_share_transfer alice bob thinkedge450)
1 : (risk__access_control)
```

List. 8: Risk Data Leak (plan)

The difference between the planner output in the case of the red team vs. in the case of the blue team reflects the asymmetric nature of cyber warfare where the defender needs to do everything right; in contrast, the attacker only needs to succeed once.

5 Related Work

There is no shortage of literature that focus on specialized tasks from the red team point of view: Automate attacks against networks with hundreds of machines Boddy et al. [Bo05], Obes et al. [OSR13] and Ron Alford et al. [Ro22]. From the blue team point of view: Role-based access control optimization Benedetti; Mori [BM19].

A successful attempt[11] has been made to port *Fast Downward* to WebAssembly (`WASM`) using Emscripten by Tran et al. [Tr20] with a small performance penalty[12]. This significantly reduces the need for server-side setup, facilitates integration with other modern web-based tools, and improves data privacy considerably as the planning process and output never leave the user's browser.

6 Summary

The proposed approach is flexible enough to be utilized by red and blue teams to secure industrial automation and control systems (IACS) and comply with the most recent version of any particular standard. Red teams can uncover potential threats on an organizational level, and blue teams check a high-level guide before introducing a new system asset to the environment.

[11] `https://project.cispa.io/fd-in-browser/`
[12] depending on the web browser

Literatur

[20] IEC 63096; Nuclear Power Plants – Instrumentation, Control and Electrical Power Systems Security Controls. International Electrotechnical Committee, 2020, ISBN: 978-2-8322-8837-5.

[BM19] Benedetti, M.; Mori, M.: On the Use of Max-SAT and PDDL in RBAC Maintenance. Cybersecurity 2/1, S. 19, Dez. 2019, ISSN: 2523-3246.

[Bo05] Boddy, M.; Gohde, J.; Haigh, T.; Harp, S.: Course of Action Generation for Cyber Security Using Classical Planning./, S. 10, 2005.

[BQB17] Bochtler, J.; Quinn, E. L.; Bajramovic, E.: Development of a New IEC Standard on Cybersecurity Controls for I&C in Nuclear Power Plants – IEC 63096./, S. 11, Juni 2017.

[D 98] D. McDermott; M. Ghallab; A. Howe; Craig A. Knoblock; A. Ram; M. Veloso; Daniel S. Weld; D. Wilkins: PDDL - The Planning Domain Definition Language./ , 1998.

[EH04] Edelkamp, S.; Hoffmann, J.: PDDL2. 2: The Language for the Classical Part of the 4th International Planning Competition./Technical Report 195, University of Freiburg, Jan. 2004.

[He06] Helmert, M.: The Fast Downward Planning System. Journal of Artificial Intelligence Research 26/, S. 191–246, Juli 2006, ISSN: 1076-9757, arXiv: 1109.6051 [cs].

[HLF04] Howey, R.; Long, D.; Fox, M.: VAL: Automatic Plan Validation, Continuous Effects and Mixed Initiative Planning Using PDDL. In: 16th IEEE International Conference on Tools with Artificial Intelligence. IEEE Comput. Soc, Boca Raton, FL, USA, S. 294–301, 2004, ISBN: 978-0-7695-2236-4.

[NM19] Nebel, B.; Mattmüller, R.: Principles of AI Planning - PDDL, Albert-Ludwigs-Universität Freiburg, Nov. 2019.

[OSR13] Obes, J. L.; Sarraute, C.; Richarte, G.: Attack Planning in the Real World./, S. 8, Juni 2013.

[Ro22] Ron Alford; Lukas Chrpa; Mauro Vallati; Andy Applebaum: Knowledge Reformulation and Deception as a Defense Against Automated Cyber Adversaries. Proceedings of the ... International Florida Artificial Intelligence Research Society Conference 35/, Mai 2022.

[Te18] Tellabi, A.; Zid, I. B.; Bajramovic, E.; Waedt, K.: Safety, Cybersecurity and Interoperability Aspects in Modern Nuclear Power Plants./, S. 12, 2018.

[Tr20] Tran, N.; Speicher, P.; Kuennemann, R.; Backes, M.; Hoffmann, J.; Torralba, A.: Planning in the Browser./, S. 3, 2020.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1539

Model-based Integrity Monitoring of Industrial Automation And Control Systems

Ludger Peters[1] Mahmoud Khalaf[2] Karl Waedt[3] Josef Schindler[4] Siwar Belaidi[5]

Abstract: This paper aims to enhance cyber security within Electrical Power Systems (EPS) of power plants by extending and using an updated plant simulator. In this paper, we assume a sophisticated attacker, as part of an Advanced Persistent Threat (APT), who gradually damages or manipulates primary assets (in the sense of ISO/IEC 27005:2018, e. g. main cooling water pumps, feedwater pumps, safety valves, and circuit breakers). Accordingly, we assume that the attack agent performs gradual manipulations at the application level. Detecting and predicting a potential anomaly is designed and implemented based on machine learning of expected behavior. The paper will include examples of attacks executed over an extended time period by gradually manipulating combinations of analog and binary signal values or set-points. Challenges related to the training of the detection algorithms, avoidance of false positives, and concise reporting to non-security domain experts will also be addressed.

Keywords: Digitial Twin; Machine Learning (ML); Deep Learning; Cyber Security; IIoT; Cyber-Physical System (CPS); Security Controls; Industrial Automation and Control System (IACS); Electrical Power System (EPS)

1 Introduction

Digital control systems have progressed significantly in the last decades and are present in nearly all Industrial Automation and Control System (IACS). With the current iterations, based on standard Information Technology (IT) hardware, old Industrial Control System (ICS) operating in a closed and isolated environment are being replaced and integrated into a factory-wide IACS. However, in the most critical part of Nuclear Power Plant (NPP), the nuclear containment, including the primary circuit and the related ICS, only outgoing connections will be allowed in the foreseeable future.

IACS have evolved into a collection of ICS, communication technology, and other hardware which work together to automate and operate industrial processes. While those systems offer significant benefits to optimize the control of entire Supply Chains, they bear the risk of breaching security. The most relevant example of such vulnerability of an IACS is the

[1] Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany lpeters@informatik.uni-freiburg.de
[2] Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Germany khalaf@cs.uni-freiburg.de
[3] Framatome GmbH, Erlangen, Germany Karl.Waedt@framatome.com
[4] Friedrich-Alexander University of Erlangen-Nuremberg josef.schindler@covalion.net
[5] Framatome GmbH, Erlangen, Germany siwar.belaidi@framatome.com

doi:10.18420/inf2022_132

Stuxnet attack that was executed in June 2010. “Stuxnet” attacked the control system at the Iranian nuclear power station, where attackers gained access to the control systems and destroyed over 1000 centrifuges. Before “Stuxnet”, the Davis-Besse NPP was attacked in 2003 and the Hatch NPP in 2008. [Pe18]

This paper targets the enhancement of cyber security within nuclear power plants, modeling one assumed attack scenario. It continues the research Ms. Gupta started within her Ph.D. Thesis [Gu21]. The topic of this paper lies within two independent research projects. The first project is a cooperation of the Framatome company with different universities like the University of Siegen, the University of Erlangen, and the Magdeburg-Stendal University of Applied Sciences. This first project targets the enhancement of cyber security within nuclear power plants. The second project is the CRP J008 project organized by the International Atomic Energy Agency (IAEA). Within this project, the Asherah Nuclear Simulator (ANS) is developed to support the security research on NPP. [Gu21].

While the paper focuses on an NPP example, the approach can be generalized for other industry scenarios, assuming that the interaction between primary and supporting assets is well formalized.

2 Related Research and Technologies

As mentioned above, this chapter outlines the foundations for the main body of this paper. The explained research and technologies were considered during the development of the prototype. They directly or indirectly guided the choices and decisions relevant to the development of the demonstrator.

2.1 Behaviour Anomaly Detection (BAD)

Behaviour Anomaly Detection (BAD) describes the detection of unexpected action within Cyber-Physical System (CPS). Any system should react to the input variables with specific, predictable behavior. A behavioral anomaly occurs if the reaction differs from the expected actions.

In recent years multiple research studies have been executed on different possibilities of the automated detection of behavior anomalies within electronic systems. Several papers were published recently for the unique field of anomaly detection in CPS. Most of this paper will be guided by the following three publications, which the analyzed physical systems will generally separate.

Both Ramen et al. [RSM20] and Xu et al. [XAY21] used the water treatment system [Au20; Ta18], designed for CPS security analysis to test their approaches. In opposite to the dynamic behavior of NPP, which ideally operates at the maximum energy output with a constant state

in the subsystems, the water treatment testbed operates in cycles. The two main approaches to behavior detection in research are explained in the following paragraphs.

MLP-CUSUM based Detection Raman et al. [RSM20] explore the possibility of using a combination of an Multilayer Perceptron (MLP) to predict the behavior of the system and a cumulative sum (CUSUM) to compare the prediction to the actual system states.

Digital Twin based Detection Xu et al. [XAY21] explored another approach to detect attacks within the same system. The resulting approach is called Anomaly deTection with digiTAl twIN (ATTAIN). The observed CPS is mirrored continuously into a digital twin within this approach.

The general idea of this approach is to use the built digital twin model to indicate if the CPS is operating normally. This allows their ATTAIN approach to use the recorded real-time data to optimize the digital twin in parallel to detect anomalies.

2.2 Asherah Nuclear Simulator part of the IAEA CRP J008 Project

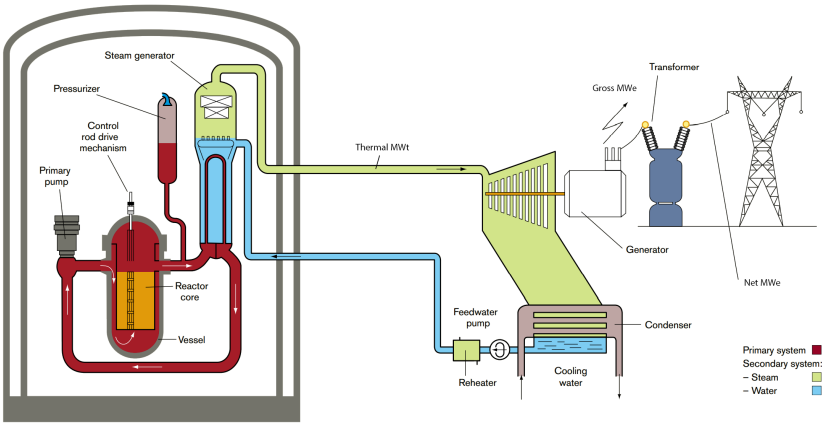


Abb. 1: Pressurized Water Reactor [Ar09]

The Asherah Nuclear Simulator (ANS) is a model of a nearly 3 GW thermal two-loop Pressurized Water Reactor (PWR). This model is loosely based on the Three Mile Island (TMI) reactor design, as installed in Harrisburg (Pennsylvania, US). Nearly all current reactor designs use a four-loop design. For the model, a two-loop design is more straightforward but sufficient for simulation as it still maintains the relevant characteristics. The model includes all relevant equipment for the primary and secondary cycles. Figure 1 shows the general layout of a PWR-based NPP, to give a general understanding of the component locations. [Bu20; Bu21; Gu21; Sh19]

2.3 Detection Duration

As developed in this paper, the demonstrator evaluated the additional information gained to comprehensively detect, respond and fully contain a cyber attack. The problem is that depending on the research study, the current time frame of detection varies significantly. Most organizations claim that a cyber-attack should be contained in a $1 : 10 : 60$ scheme [Cr22]. This means that a cyber-attack should be detected within 1 minute, investigated within 10 minutes and contained within 60 minutes. According to an analysis by CrowdStrike, it currently takes about seven days, working 24 hours straight to fully contain a cyber-attack on average. The IBM Security group even states that this work takes about 280 days on average in their 2020 report on the cost of data breaches [IB20].

3 System Settings

3.1 Attack Description

All attacks in this paper consider the direct manipulation of control systems. These attacks are also expected to result from an Advanced Persistant Threat (APT). Although all necessary levels of security are implemented in NPPs, attackers can overcome these. At this point, the impact of these attacks is devastating and will result in a short or long time hardware damage. Especially in a larger attack surrounding, the power grid's stability could be impacted.

Figure 2 displays the assumed attack flow, in which the attacker gains undetected access to the maintenance network of the NPP. This access allows the manipulation of different ICS, which must be detected to prevent further damage.

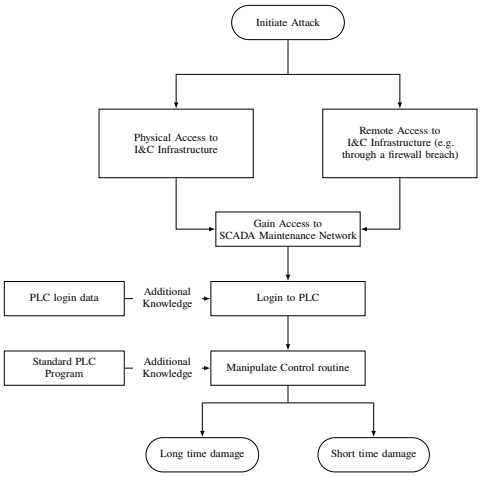


Abb. 2: General Attack Flow

3.2 System Threat

3.2.1 Scenario

This scenario focuses on an attack on the control system itself. It is assumed that the attacker wants to damage the plant and can change the programming of the control system itself. With this access, he can manipulate single control variables. He also can manipulate the Human Machine Interface (HMI) data so that the operating personnel cannot detect his changes.

Impact In this scenario, the attack impacts several layers of the safety measures applied inside a nuclear power plant. Consequently, the attack might directly or indirectly lead to an automatically reduced power output or even a reactor emergency shutdown.

Assume that the feed water pump speed is iterated between 75% and 125% of the nominal pump speed, but the correct nominal speed at 100% is displayed to the operators. This invisible manipulation will lead to increased wear on the components and problems with the heat transport between the steam generator and the turbine. This attack has different possible effects on the operation of the NPP. This might be missing water in the feed water reserve tank or a underfill in the steam generator, which can potentially cause a full plant shutdown. The detection of this scenario is tested in section 5.3.2.

3.3 System Choice

Within this section, basic choices for the detection system are described. The system will be designed to detect the attacks before any component is damaged. At first, the decision between direct or indirect BAD is explained.

3.3.1 Detection Background

Section 2.1 is referencing three papers regarding BAD within different surroundings. The authors of the first paper [RSM20] describe the detection within a Water Treatment system built for cyber security analysis. The authors of the second paper [XAY21] use the same system and apply the combination of a Timed Automaton Machine (TAM) and a Generative Adversarial Network (GAN) to detect anomalies within the system.

These approaches commonly use logged data of a correct working system to build a prediction model. Additionally, all the authors split their detection model into two parts: the prediction and the detection.

3.3.2 Direct Detection vs. Prediction based Detection

Direct Detection means that the Machine Learning (ML) algorithm detects that an attack is executed on the system. The algorithm needs to be trained with the system states as an input and, in the most basic form, a label if an attack is executed on this input patch as an output label. This leads to general questions about the whereabouts of the attack. This includes the timeframe, the impact, and the affected systems.

To analyze these questions, a deep dive into the basic algorithm is necessary. An ML-based prediction algorithm in this surrounding needs to be trained on labeled data or needs to learn in an attack surrounding. Both options are impossible as we can neither predict nor execute all possible attacks.

Therefore, we need another option where the decision does not take place in an ML algorithm's "black box".

With a prediction-based approach, the detection is split into two tasks. In the first task, the expected system states are predicted. In the second task, the expected system states are compared to the measured system states.

Due to the impossibility of implementing a direct detection algorithm, this paper will use a prediction-based approach.

3.3.3 ML vs. Digital Twin based Prediction

As outlined in section 2.1, both an ML-based and a digital twin-based prediction are suitable options for predicting the system states.

This paper emphasizes on using an MLP based prediction as proposed by Raman et al. in [RSM20]. The advantage of this approach is the higher flexibility, allowing for changing the Programmable logic controller (PLC) configuration without much effort as the MLP model can just be retrained with the newly recorded data.

A digital twin is recommended in a physical system, as these systems will operate over a long time horizon, with subsystems being replaced several times in the expected lifetime.

4 Implementation of the detection system

The Implementation of the system is split into three separate programs.

The first program is used to collect the initial log data of the system. As the overall system will use the PLCs provided by the ANS to demonstrate the detection capabilities of the system, the data is accessed by using the provided Open Platform Communications Unified

Architecture (OPC UA) server using the asyncua framework [Fr22]. The data is logged into a `*.csv` file by the python logging system configured according to the `*.csv` specifications.

The second program is responsible for finding and training the ML model. This paper is based on the combination of a TensorFlow-based model [De15] and the KerasTuner [OM19] for the optimization of the model layout. This model predictions the next timestep of the system based on several input timesteps.

The third program is the detection system which uses the same framework as the first program to access the sensor data on the OPC UA server. As soon as enough data is collected to start the prediction (the available number of timesteps exceeds the number of timesteps required by the model) the detection process is started.

5 System Testing

5.1 Definition of Experiments

For comparison, all experiments regarding the same goal should also follow the same rules. In the first experimental step, the plausibility of model training is tested. Afterwards the reaction of the system to the different attack scenarios is test.

Two different runtime scenarios will be used for testing. The simulator will run at full power for about 17 minutes in the first one. The second test will run for 55 min and include several defined power changes which are repeated in every run. This is used to simulate a load-following operation, thus, it is to be stated that kind of operation is not preferred within nuclear power plants. However, due to current regulations, the reactor must be capable of this kind of operation [Mü03].

5.2 Comparison Data

This test is executed to evaluate the general functionality of the final prediction algorithm.

The result of these runs for the feed water tank level is shown in figure 3. The other information needed is the scale of these variables, which is included in table 1. These figures show that the prediction system generally works for both simulation runs. However they also show that the data from the load simulation run is necessary to create a tolerance reference to avoid false positives.

5.3 Attack Scenarios

Due to the construction and design of the current NPPs, the scenarios will focus on the secondary system. In the current design, the nuclear part of the control systems are also called

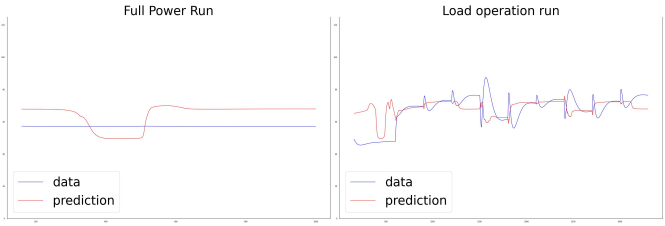


Abb. 3: Feed Water Tank Level on Databroker Comparison Run

"Nuclear Island". The data communication with these ICS is only allowed one-directional, as the communication direction is from the ICS controlling the nuclear components to the outside world.

The ICS of high consequence facilities, like NPPs, which require the highest level of security are only connected to the IACS of this facility via secure, deterministic, unidirectional data communication pathways as per the Defensive Computer Security Architectures (DCSA) specification. [IN21]

The demonstrator will be tested on the ICS controlling the secondary circuit. With the results transferable to the primary ICS as only the data source and manipulation point change.

Within these scenarios, two variables will be analyzed deeper as the system detects an anomaly in nearly all scenarios—the exception changes, where the result of the operability is unclear. Two selected variables will be plotted for all attacks. Table 1 displays all variables, their corresponding scale, and the tolerance data resulting from the run in section 5.2.

Tab. 1: Tolerance Data for selected Variables

		Scale		Tolerance (Scale ≡ Variable amount)	
Sensor	Unit	0 ≡	100 ≡	lower	upper
FW_TankLevel	m	2.6	5.0	$-28 \equiv -0.67$	$30 \equiv 0.72$
CD_Level	m	0.6	1.3	$-27 \equiv -0.2$	$20 \equiv 0.14$
GN_GenElecPow	W	0.0	919×10^6	$-41 \equiv -3.8 \times 10^8$	$28 \equiv 2.6 \times 10^8$
RX_ReactorPress	Pa	14×10^6	15×10^6	$-51 \equiv -5.1 \times 10^5$	$12 \equiv 1.2 \times 10^5$
SG_Level	m	13.3	16.4	$-8 \equiv -0.25$	$20 \equiv 0.62$

5.3.1 Value Spoofing

Assumptions The steam generator is responsible for exchanging heat between the primary and secondary cycle. In this attack, the water level in the steam generator is set to 0.15 m and 1 m below the target level in the 100% run and 0.15 m in the load simulation run. This will lead to a higher water level in the steam generators and a lower level in the feed

water tank. This attack does impact the heat transfer in the steam generator and will lead to increased wear within the primary circuit.

Detection In the 100% run with 14 *m* as a steam generator level spoof, the first alarm is raised 22 *s* after the attack started on the steam generator level being too low. 1 *s* later, an alarm is also raised on the feed water tank levels being too high. The critical difference to the 14.85 *m* spoof is that the alarms for both the steam generator level and the feed water tank level are not raised in this run. The other alarms are raised about 1 *min* earlier, which indicates a higher impact on the heat transfer within the simulation.

For the second test with the steam generator spoofed to 14.85 *m* and the load simulation being executed, the first alarm is raised 5 *s* after the attack started, also on the feed water tank level being too low. The alarm on the steam generator level is raised 3.5 *min* after the attack has started. The other alarms, raised about 45 *s* later, are regarding the condenser and the condensate extraction pump.

5.3.2 Value Iteration

This attack scenario focuses on the possibility of overloading the feed water pumps. This attack scenario has been mentioned in chapter 3.

Assumptions The feed water pumps are responsible for the water flow within the secondary cycle. If the attacker can increase the wear of these components, they might fail within the normal operation. This would lead to a complete reactor shut down as the water transport from the core to the turbine is damaged. To test the detection's basic capability, the pump speed is iterated between ±25% again for a full power run in the first test and the load simulation run in the second test.

Detection In the 100% run, the first alarm is raised 25 *s* after the attack started on the steam generator level being too low. 52 *s* later, an alarm is also raised on the feed water tank levels being too high. Between 2 and 3.5 *min* after the start of the attack, several other alarms are raised.

The first alarm within the load simulation test is raised 2 *min*, and 13 *s* after the attack has started on the water and the steam temperature in the condenser being too high. In the following 23 *min*, several other errors are thrown.

5.4 Test Results

These results correspond to the result from the papers mentioned in section 2.1. The evaluation of the detection time of this algorithm is complex due to several related and open questions. As described in section 3.1, this algorithm is designed to prevent hardware damage in case of a successful attack. The results show larger manipulations are detected within seconds when the manipulation is more extensive than the detection tolerance. More data regarding the manipulation impact on the components' lifetime are necessary to evaluate the sufficiency of the tolerance area.

6 Conclusion

This paper aimed to find a solution to detect the manipulation of attackers that surpassed the protective security controls of the ICS of NPPs, within the context of the SMARTEST2 R&D project. As stated in section 3.1, these attacks are highly unlikely, but if they occur, they have a critical impact on the operation of the attacked system. The referenced literature in section 2.1 indicated that two general approaches are possible within a periodically dynamic system. Conversely, the target IACS typically is a stable system that can follow the currently needed power output. This fact is used to simulate multiple power loads and use them for the generation of the training data. The result of the experiments in chapter 5 proved that the chosen system setup is a valid answer to the main objective of finding anomalies within the IACS of NPPs.

Model and Prediction The system produces more than adequate results given the mentioned constraints, and false positives can be eliminated by choosing the corresponding tolerance values. However, to achieve this elimination, it is necessary to choose the tolerance starting with shortly over 10% up to nearly 60% within the scaled data, compare table 1. These tolerance levels should be lowered drastically in the future. One option is the realization through the reconfiguration of the model's search space by increasing the number of possible neurons in each layer substantially. General research on the topic of network optimization is provided by the AutoML Group [Ma22] The other option is testing the second approach mentioned in section 3.3.3, especially in the form summarized in section 2.1 for the digital twin.

The following paragraphs summarize the possible future improvements and additional work.

Notification of the Operators As explained above, choosing the right tolerance area can nearly avoid all false positives. However, that does not answer how the operators are notified. The suggested method is to use the available OPC UA server and create a new subspace for the alarms. The corresponding fault direction is then set as an alarm for the detected

value anomaly and additionally selected if there is an iteration between the value being too high and too low. It is proposed to reset these alarms by the operators so that short-time anomalies can be researched. It is not proposed to pass through the amount of the anomaly, as the tolerances are too high to give a conclusion at the moment.

Impact and Reaction Time As detailed in the test result in section 5.4, further research is necessary to evaluate the necessary tolerance level and reaction time.

Data Accuracy Another critical point within the deployment of such a system is data reliability. The sensor data needs to be trustworthy and correct to ensure a reliable prediction. Additional research needs to be executed to establish a trustworthy system.

Literatur

[Ar09] Areva GmbH: The Path of Greatest Certainty, 2009, URL: http://www.epr-reactor.co.uk/ssmod/liblocal/docs/EPR%20Interactive/Brochures/300709_EPR_52pages.pdf, Stand: 05. 06. 2022.

[Au20] Aung, Y. L.; Tiang, H. H.; Wijaya, H.; Ochoa, M.; Zhou, J.: Scalable VPN-forwarded Honeypots: Dataset and Threat Intelligence Insights. In: Sixth Annual Industrial Control System Security (ICSS) Workshop. ACM, Austin TX USA, S. 21–30, Dez. 2020, ISBN: 978-1-4503-9002-6.

[Bu20] Busquim e Silva, R.; Shirvan, K.; Cruz, J.; Marques, R.; Marques, A.; Piqueira, J.: Advanced Method for Neutronics and System Code Coupling RELAP, PARCS, and MATLAB for Instrumentation and Control Assessment. Annals of Nuclear Energy 140/, S. 107098, Juni 2020, ISSN: 03064549.

[Bu21] Busquim e Silva, R.; Piqueira, J.; Cruz, J.; Marques, R.: Cybersecurity Assessment Framework for Digital Interface Between Safety and Security at Nuclear Power Plants. International Journal of Critical Infrastructure Protection 34/, S. 100453, Sep. 2021, ISSN: 18745482.

[Cr22] CrowdStrike: Report: Organizations Take 162 Hours to Detect and Respond to Breaches, 2022, URL: https://www.crowdstrike.com/press-releases/crowdstrike-study-finds-organizations-take-over-162-hours-to-detect-respond-to-breaches/, Stand: 28. 05. 2022.

[De15] Developers, T.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, Zenodo, Software available from tensorflow.org, Mai 2015, URL: https://www.tensorflow.org/.

[Fr22] FreeOpcUa: Opcua-Asyncio, Free OPC-UA Library, Juni 2022, URL: https://github.com/FreeOpcUa/opcua-asyncio, Stand: 12. 06. 2022.

[Gu21] Gupta, D.: Nuclear Safety Related Cybersecurity Impact Analysis and Security Posture Monitoring, Diss., 2021.

[IB20] IBM Security: Cost of a Data Breach Report 2020./, S. 82, 2020, URL: https://www.ibm.com/security/digital-assets/cost-data-breach-report/1Cost%20of%20a%20Data%20Breach%20Report%202020.pdf, Stand: 28. 05. 2022.

[IN21] INTERNATIONAL ATOMIC ENERGY AGENCY: Computer Security Aspects of Design for Instrumentation and Control Systems at Nuclear Power Plants. INTL ATOMIC ENERGY AGENCY, S.l., 2021, ISBN: 978-92-0-109020-1.

[Ma22] Marius Lindauer, F. H.: AutoML | AutoML, 2022, URL: https://www.automl.org/automl/, Stand: 12. 06. 2022.

[Mü03] Müller, K.: Lastfolgebetrieb und Primärregelung./, S. 13, 2003, URL: https://www.ktg.org/ktg-wAssets/docs/fg-bet-rph-lastfolgebetrieb-kki2.pdf, Stand: 09. 05. 2022.

[OM19] O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L. et al.: KerasTuner, 2019, URL: https://github.com/keras-team/keras-tuner.

[Pe18] Peters, L.: Prototype Implementation of a Virtualized Instrumentation and Control System for Nuclear Power Plants./, S. 86, Sep. 2018.

[RSM20] Raman MR, G.; Somu, N.; Mathur, A.: A Multilayer Perceptron Model for Anomaly Detection in Water Treatment Plants. International Journal of Critical Infrastructure Protection 31/, S. 100393, Dez. 2020, ISSN: 18745482.

[Sh19] Shirvan, K.; Busquim e Silva, Rodney; Piqueira, José Roberto Castilho; Marques, Ricardo Paulino: Development of the Ahserah Nuclear Power Plant Simulator for Cyber Securrity Assessment./, S. 10, 2019.

[Ta18] Taormina, R.; Galelli, S.; Tippenhauer, N. O.; Salomons, E.; Ostfeld, A.; Eliades, D. G.; Aghashahi, M.; Sundararajan, R.; Pourahmadi, M.; Banks, M. K.; Brentan, B. M.; Campbell, E.; Lima, G.; Manzi, D.; Ayala-Cabrera, D.; Herrera, M.; Montalvo, I.; Izquierdo, J.; Luvizotto, E.; Chandy, S. E.; Rasekh, A.; Barker, Z. A.; Campbell, B.; Shafiee, M. E.; Giacomoni, M.; Gatsis, N.; Taha, A.; Abokifa, A. A.; Haddad, K.; Lo, C. S.; Biswas, P.; Pasha, M. F. K.; Kc, B.; Somasundaram, S. L.; Housh, M.; Ohar, Z.: Battle of the Attack Detection Algorithms: Disclosing Cyber Attacks on Water Distribution Networks. Journal of Water Resources Planning and Management 144/8, S. 04018048, Aug. 2018, ISSN: 0733-9496, 1943-5452.

[XAY21] Xu, Q.; Ali, S.; Yue, T.: Digital Twin-based Anomaly Detection in Cyber-physical Systems. In: 2021 14th IEEE Conference on Software Testing, Verification and Validation (ICST). IEEE, Porto de Galinhas, Brazil, S. 205–216, Apr. 2021, ISBN: 978-1-72816-836-4.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1551

Securing javascript runtime of OPC UA deployments

Josef Schindler [1], Siwar Belaidi [2], Erkin Kirdan [3], Dr. Karl Waedt [1]

Abstract: OPC UA is a commonly used machine-to-machine communication protocol. As it is used in industrial critical infrastructure, its security is vital. OPC UA has several proprietary and open-source implementations in different programming languages. Among them, a JavaScript implementation node-opcua stands out due to advantages such as ease of use, code maturity etc.

JavaScript is a just-in-time compiled programming language that is executed by a runtime system. Node.js is the most common runtime environment to run JavaScript programs. Deno is it a potential and unofficial successor as it is developed by the original author of Node.js. One promised improvement of Deno is the focus on higher security, such as having a restricted file system and network access by default.

Since (1) not all libraries have been adopted yet for for Deno, since (2) this new runtime comes with a compatibility mode featured first in version v1.15, since (3) there is very few research literature yet and since (4) there is no current dedicated OPC UA implementation for Deno so far, we present and evaluate measures to overcome errors when running node-opcua on Deno in this paper.

Keywords: opc ua; security; node.js; deno; javascript

1 Introduction

In industry, OPC UA is used as a communication protocol between field devices and devices to the cloud. One of the greatest strengths is its interoperability, i.e., not caring which OPC UA stack is communicating to one another. This results in many implementations, also open-source, like the node-opcua stack. It receives good feedback from the community due to its simple coding. A newer runtime for JavaScript is Deno, which promises even higher security, and is partly able to run Node.js code.

The paper is structured as follows. In the next section we present specifics of OPC UA, Node.js and Deno, and provide their differences. In section 3, we then provide some workarounds that we found trying to run node-opcua on Deno. Finally, section 4 depicts the conclusion.

[1] Framatome GmbH, ICETA-G Department, Paul-Gossen-Straße 100, 91052, Erlangen, Germany firstname.lastname@framatome.com
[2] Framatome GmbH, ICETA-G Department, Paul-Gossen-Straße 100, 91052, Erlangen, Germany siwar.belaidi@framatome.com
[3] Technical University Munich erkin.kirdan@tum.de

doi:10.18420/inf2022_133

2 Background and related work

2.1 OPC UA

OPC UA is a platform-independent service oriented protocol commonly used for machine to machine communications. Fields of application includes Control Systems, Energy and utilities, Manufacturing Execution Systems and Enterprise Resource Planning Systems. OPC UA is a client-service protocol which provides a consistent and integrated AddressSpace and service model. OPC UA AddressSpace allow servers to provide clients with an object model represented as a set of variables and methods. Services are actions offered by servers to clients. Discovery service is an example of operation offered by OPC UA servers, it allows the clients to discover the endpoints implemented by the server. An other example of services, the Securechannel service that allows clients to establish a communication channel with the server [OP20a, OP20b, OP20c].

OPC UA relies on SecureChannels and Sessions. The OPS UA client start opening a SecureChannel to ensure integrity, confidentiality and application authenticity. A Session is then created between the client and the server on top of the SecureChannel for user authorization and authentication. The client can finally access data or call methods on the server through Services [Mu21].

OPC UA has several open-source and commercial implementations. Among them, the node-opcua implementation fully written in Javascript and Typescript.

2.2 Node.js

Node.js is an open source runtime environment for Javascript powered by the V8 engine, the core of Google Chrome. Node.js is built on non-blocking asynchronous I/O paradigms, allowing it to hundle thousands of concurrent tasks on a single server and making it ideal for highly scalable, data intensive, and real-time web and mobile applications. Node.js uses npm as a default packages manager, a command-line utility that allows packages installation, versions and dependencies management, npm also provides an online repository that hosts open source packages for Node.js [Op22].

2.3 Deno

Deno is a modern runtime environment for Javascript and Typscript built in Rust programming language and based on the V8 Google engine. It was created by Ryan Dahl the original author of Node.js. Deno is designed to focus on higher security, it provides secure defaults by restricting files and network access. Unlike Node.js, Deno does not use npm for resources fetching, it uses modules referenced as URLs or file paths, it also provides a list of standard modules hosted on an online repository and distributed via URLs [Da22].

2.4 Comparison of Node.js and Deno

Deno provides secure defaults - unless specified, file, network and environment access are not allowed. Access to security sensitive functionalities requires a permission to be granted to an executing script through command line flags or a runtime permission prompt [Da22].

Deno also provides built-in tooling to improve developer experience such as unit testing, code formatting, and linting. Unit testing is a software testing method in which small units of source code are tested as applications are developed. Unit testing allows to detect defects at an early stage of the development process, these defects can be remediated faster and with lower cost than at a later security hardening stage. It also allows to validate the proper implementation of code changes required by secure code reviews. By using unit testing in conjunction with secure code reviews, developers can identify vulnerabilities, validate security functionalities of the code and verify whether these functionalities are able to mitigate the identified security defects. Linting provides code reviews by analyzing source code and detecting potential syntax errors and deviations from coding styling rules. It allows to improve code and maintain its consistency and readability. Linting can be used as a complementary approach to unit testing to build strong test scenarios that cover all parts of the code.

Exceptions or errors in computers during the execution of programs are defined by anomalous or exceptional conditions that occur. Indeed avoiding them 100% is desired, but in real applications this is not possible and therefore needs adequate handling. This so called exception or error handling is crucial to create a secure framework and secure applications by avoiding non-determined system state and behavior after their occurrence. Exemplary, Wu et al. [Wu17] found many vulnerabilities in Android code, which resided from a poor and complex exception handling in Android.

Probably, one big difference is that Deno always dies immediately on uncaught promises, where Node.js currently handles rejections by emitting a deprecation warning to stderr. Right now, Node.js is running a survey to better understand the way users are dealing with this. As explained above, uncaught errors may lead to vulnerabilities. They are ok most of the time, but apparently Deno does not want to ignore them, forcing developers to write more determined code.

3 Workarounds & Evaluation

In this section, the code examples are adopted from a basic server implementation with node-opcua library [St22]. The schematic debugging process is in line with the examples outputs when processing the code and is depicted in general in figure 1. It shows the stages of the debugging process on the left side. The arrows in between the stages point out possible directions. On the right side, belonging, alternative measures are presented. The circular arrows depict weather measures for the stages 2 to 4 may occur several or just a single time.

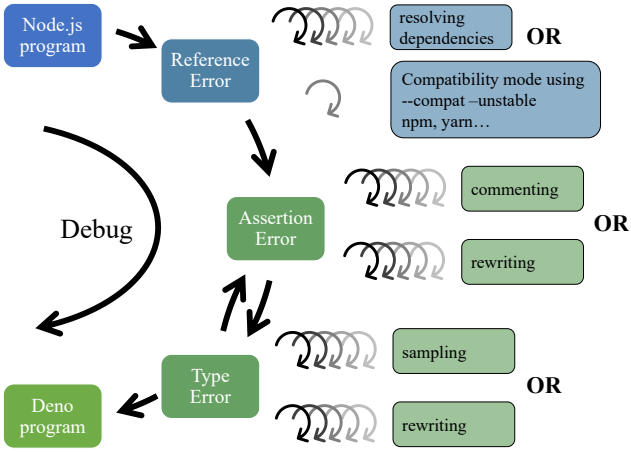


Fig. 1: debugging process for a basic server of node-opcua running with Deno.

Running a Node.js designed program with Deno without adjusting the code will presumably result in

```
error: Uncaught ReferenceError: require is not defined
```

This error occurs as the syntax of library imports changes from Node.js to Deno along with the paradigm of managing those dependencies as explained in section 2.2 and 2.3. There are two workarounds:

1. changing and finding the correct equivalent for Deno: This is the most adequate solution. But on the backside, not every library is supported by Deno yet, and a programmer has to go through every `require` statement.
2. enable the compatibility mode using `--compat`, first introduced in Deno v1.15: Besides allowing for Node.js import syntax, this mode provides global Node.js variables and support for the Node.js resolution algorithm [Da22]. Note, `--compat` option requires to download the Node.js dependencies with npm, yarn or similar, and a further flag `--unstable`, which can lead to unsecure code.

The next type of errors will be uncaught promises:

```
error: Uncaught in promise AssertionError
error: Uncaught in promise TypeError: dns.lookupService is not a function
error: Uncaught in promise TypeError: child_process.exec is not a function
```

Deno treats uncaught promises very differently compared to Node.js, i.e. will terminate the execution occurrence immediately. Node.js is less strict as described in section 2.4 and thus less secure - passing uncaught promises can lead to unexpected and therefore insecure states. Wu et al. [Wu17] present an example that can exploit such insecure behavior.

In our scenario, depending on what the reason for the uncaught promise is, three fixes silenced or resolved these errors:

1. Commenting: Whenever the Node.js program consciously raises an assertion, one can comment the code line to mimic the behavior of Node.js to continue the execution.
2. Sampling: Here, the error can be mitigated inside functions that return simple data. Like, in the error code above `TypeError: dns.lookupService`. For a simple server that only runs locally, one can derive the output manually and sample it during the function call `callback(null, 'localhost ssh');`.
3. Rewriting: Indeed, this requires the most effort and needs a proper understanding of what the program does at specific code lines. On the other hand, this will lead to the best results even if one engineers more complex programs

4 Conclusion

This short paper provided some workarounds and debug schemes to run a Node.js designed program on Deno.

There are drawbacks in making Node.js designed programs compatible to Deno. First getting rid of package managers is no longer valid. Second, using unstable option is required. Third, one needs to handle or comment assertions. Handling takes much resources during the debugging and commenting will decrease the benefits from Deno.

However, being forced to take care of uncaught promises is one of the strong security arguments opting for Deno. Also, being able to run Node.js code might attract more coders and thus balance the strongest argument for Node.js - the mature code base.

References

[Da22] Deno. `https://deno.land`, Accessed: 2022-06-27.

[Mu21] Muehlbauer, Nikolas; Kirdan, Erkin; Pahl, Marc-Oliver; Waedt, Karl: Feature-based Comparison of Open Source OPC-UA Implementations. In (Reussner, Ralf H.; Koziolek, Anne; Heinrich, Robert, eds): INFORMATIK 2020. Gesellschaft für Informatik, Bonn, pp. 367–377, 2021.

[OP20a] OPC Unified Architecture - Part 1: Overview and concepts. Standard, International Electrotechnical Commission, Geneva, CH, November 2020.

[OP20b] OPC Unified Architecture - Part 3: Address Space Model. Standard, International Electrotechnical Commission, Geneva, CH, July 2020.

[OP20c] OPC Unified Architecture - Part 4: Services. Standard, International Electrotechnical Commission, Geneva, CH, July 2020.

[Op22] NodeJS. https://nodejs.org/en/, Accessed: 2022-06-27.

[St22] NodeOPCUA. https://node-opcua.github.io/, Accessed: 2022-06-27.

[Wu17] Wu, Jingzheng; Liu, Shen; Ji, Shouling; Yang, Mutian; Luo, Tianyue; Wu, Yanjun; Wang, Yongji: Exception beyond Exception: Crashing Android System by Trapping in Üncaught Exception". In: 2017 IEEE/ACM 39th International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP). pp. 283–292, 2017.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1557

Scalable backend representation of security posture of IIoT systems

Edin Kreho[1], Roger Djeukoua [2], Timothée Guiraud [3], Karl Waedt [4]

Abstract: The focus of this paper is the scalable modelling and database representation of cybersecurity postures as part of a framework for modelling of security artefacts. A cybersecurity posture describes the current state of protection of a system. During the lifecycle phases of a power plant or manufacturing facility this includes the semi-formal database representation of all components and subsystems of automation equipment, the controlled aggregates, the related sensors and the applied security controls. The type of database choice is discussed based on the need to serve as web backend server and scalable multi-user use. The frontend part of a comprehensive security artefacts modelling framework is not in the focus of this paper.

One focus will be on the modelling related to network security artefacts as supporting assets. This will include all network devices and network endpoints with segregations between networks by physically unidirectional security gateways and firewalls. The primary assets (that are controlled or monitored) will be selected and can include pumps, pressurizers, valves, motors, circuit breakers and similar.

The important part of the modelling is to be able to represent the potential attack vectors via supporting assets up to the primary assets that may potentially be destroyed or degraded by an attack. An important aspect of the data modeling is the security grading, as present in industrial environments, but not covered by the common IT security standards (like ISO/IEC 2700x).
The current state of the art of tools and frameworks that cover a part of the intended data modelling will be outlined. The backend of a document based (e.g. MongoDB) database for modelling the relations will be presented in more detail. The backend semi-formal representation takes into account the semi-formal approach of structuring of the supporting assets, structuring of primary assets, linking between assets and association of Application Security Controls in the sense of ISO/IEC 27034-5 and ISO/IEC 27034-5-1 (for XML/JSON representation) to supporting assets.[IS17] [IS18] A web framework will be used to interact with the backend data representation. Approaches to compare different revisions of a security posture will be outlined. This will help in planning and regularly monitoring the progress of a security posture, e.g. with regard to security audit preparations.

Keywords: Cybersecurity; Industrial Automation and Control System; Industrial Internet Of things; Concurrent cybersecurity Modelling; Primary Assets; Supporting Assets; Sdocument-oriented database

[1] University Erlangen-Nürnberg, edin.kreho@fau.de
[2] Hochschule für Technik und Wirtschaft Berlin, louis-roger.tchuegoue-djeukoua@framatome.com
[3] Framatome, timothee.guiraud@framatome.com
[4] University of Freiburg, karl.waedt@framatome.com

doi:10.18420/inf2022_134

1 Introduction

The industrial internet of things (IIoT) refers to interconnected primary assets such as sensors, pumps, pressurizers, valves, motors, circuit breakers and similar devices connected together through network in industrial applications. Usually represented as layered modular architecture it is becoming a very important part of present-day industry.

Given that perspective, one of the goals of this paper is modelling related to network security artefacts. Data modeling approach is used to model data in a consistent and predictable manner in order to make it easily manageable. The use of data modeling standards is highly recommended for all projects requiring a standard means of defining and analyzing data. The key aspect in case of data modelling in this paper will be to emphasize the importance of data modelling for representation of various methods of compromising and e.g. potentially achieving unauthorized network access to launch a cyber attack. Those attack vectors are malicious and have the purpose of damaging primary assets which are necessary for the safe and undisturbed functionality of the system.

The result of this data modeling will be a semi-formal representation. The semi-formal approach of structuring the primary and supporting assets (network resources) and linking between them will be used considering association of Application Security Controls in the sense of ISO/IEC 27034-5 and ISO/IEC 27034-5-1 (for XML/JSON representation) to supporting assets. [IS17] [IS18]

The goal of the semi-formal representation is to provide an overview of the data items, relationships between them and existing constraints without revealing too many implementation details.

In this paper the importance and different ways of the scalable backend representation of cybersecurity particular approaches will also be discussed. Manufacturing facilities (e.g. power plants) will be used as an example to describe the influence and utter importance of choosing the right technologies for scalable database management system in multi-user co-development. Choosing the right technology is especially important for database, because it serves as web backend server. The choice of the database management system directly influences the ability to provide a scalable multi-user access, which will be discussed later on.

2 Purpose and Scope

Cybersecurity threats are an important part of the industrial automation and control systems, including nuclear I&C. Given the fact that requirements on safety regarding I&C apply there, consideration of cybersecurity threats are seen as important.

IIoT systems nowadays consist for the most part of digital systems including networks that are interconnected. Having that in mind security controls for the use cases of prevention, detection and correction play an important role in protection of digital systems from external and internal threats.

Purpose of this paper is to apply a semi-formal approach in structuring of the supporting assets, primary assets and linking between assets and association of Application Security Controls. [Co20]

A schematic approach which represents above mentioned assets and is showed below in Fig. 1. (tree structure example, physical protection: top level)

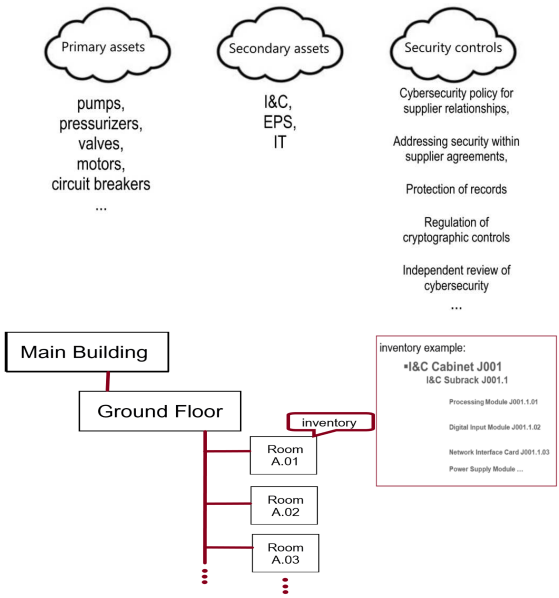


Fig. 1: Interconnection and structuring of assets representation

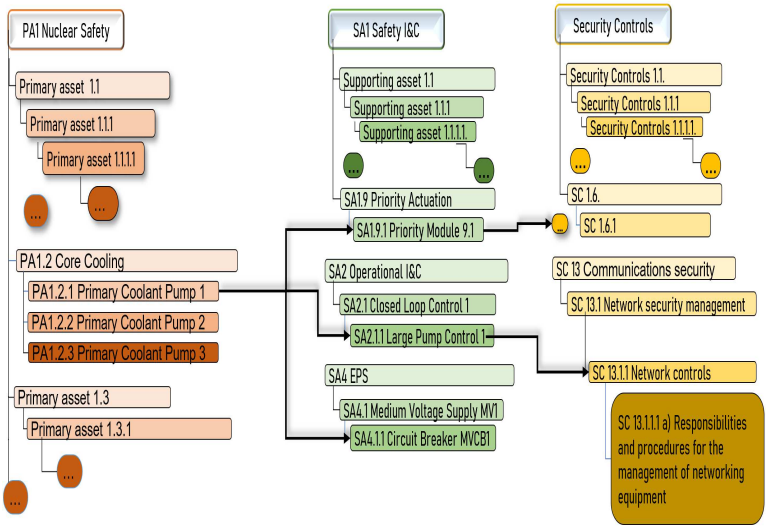


Fig. 2: Interaction among primary assets, supporting assets and security controls

In the schema above it is made clear that all assets and security controls are leaning on each other to at least some extent. Those relationships are exemplary shown in the schema. Understanding the interaction among them is also an important part to comprehensive modelling at the right level of detail. Some of the security controls and their relationships to primary and supporting assets are shown above, but there are several categories of security controls that we differentiate, such as: categories – objectives - security controls - security sub-controls.

An example of described categorization is presented in the schema: Communications security - Network security management - Network controls - a) Responsibilities and procedures for the management of networking equipment.

Without modelled Security Controls assets can be seen as obvious risks. Therefore it is of utmost importance to structure Primary and Supporting assets so that a Target Security Degree can be assigned to Primary assets and after that Security Control can be modelled.

3 Related Research and Technology Comparison

This chapter is supposed to give us a better understanding of technologies selected in order to prototypically implement the above mentioned modeling. The necessity of choosing the right technologies plays an important role when it comes to cybersecurity. Sometimes called a "solutions stack," a tech stack typically consists of programming languages, frameworks, a database, front-end tools, back-end tools, and applications connected via APIs. All of the above mentioned will be briefly explained and brought to comparison.[Ab] [Ty]

For the backend part we propose the use of Node.js with Express and MongoDB as a database. Node.js uses an event-driven, non-blocking I/O model that makes it lightweight and efficient, well-suited for data-intensive real-time applications that run across distributed devices.

Note: In an subsequent prototype Node.js may be replaced by Deno, which is was conceived by the same developers and Node.js, however with cybersecurity in mind from the very beginning. [No] [Mo]

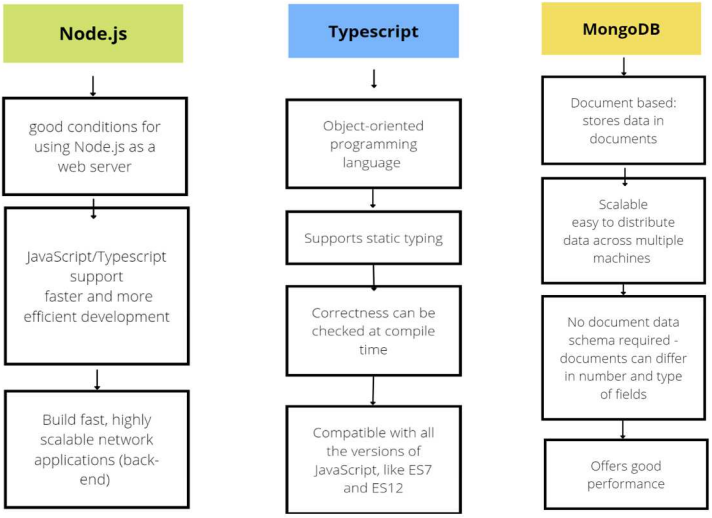


Fig. 3: Comparison among Node.js, Typescript and MongoDB

In order to better demonstrate the described importance of modelling it was decided to incorporate the functional API with the frontend. The frontend part of application uses Vue 3 as a JavaScript framework for building user interfaces. It builds on top of standard HTML, CSS and JavaScript. There are many alternatives to Vue like React and Angular. But since the focus of this topic is on modeling and semi-formal representation with need for scalability, all the above mentioned technologies for frontend are also possible alternatives.

An important part of the frontend role is interactivity with users so that the modelling part is easier to understand and visualize. The technology in mind to realize that would be a Vue 3 based component library called Element Plus, specifically speaking, e.g. the tree widget component from Element Plus component library together with further widgets for structuring, navigation and interaction. [In] [De]

As mentioned in the introduction simultaneous access by multiple users was designed and implemented on backend side. JsonWebToken in conjunction with other principles was used to provide safe Authorization and Authentication of Users.

To get an overview of the project idea and how all above mentioned technologies work together to provide such functionalities, there is a schema of functional API which is integrated with a frontend provided

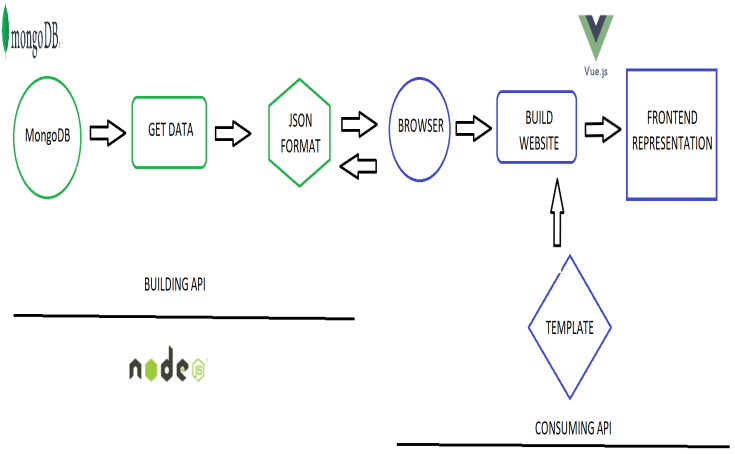


Fig. 4: API - schematic representation shows the relationships among above discussed technologies

4 Key characteristics of the scalable cybersecurity modelling

As a key characteristics the following could be denoted:

- Scalability
- Multiple Cybersecurity specialist can simultaneously edit parts of the same model
- Baseline for the definition of attack scenarios
- Baseline for performing risk assessments
- Baseline for evaluating the current security posture

Scalability is an important factor for a framework whenever there is a chance of software needing a change to meet new users needs. Having that in mind all technologies used in order to provide semi formal representation support scalability to at least some extent. This allows whole applications to adapt to new requirements when it comes to more data, users or multiple requests coming parallel from different users. Above mentioned and already discussed multi user access will provide necessary simultaneous editing support of the parts of the same model.

A cybersecurity posture describes the current state of protection of a system. It does not only include the condition of IT infrastructure, but also the human behavior, practices and processes.
They allow us to look at the big picture, to take so called holistic approach in order to establish the likelihood of a breach. This is a far more realistic approach compared to individualistic look at components, which can lead to incorrect and potentially wrong conclusions. Therefore it is important to note that given the interactive representation of assets (primary, secondary, and their connections) it is made possible to easier understand the cybersecurity posture of a system.

The structural, semi-formal and interactively changeable representation of the security artefacts will give us guidelines for possible attack scenarios, and allow us to easily visualize them. That way risk assessments are much easier to perform, and the evaluation of the current security posture is also achievable based on the observations made through modelled representation of security assets. [Wh]

References

[Ab] About Javascript, URL: https://developer.mozilla.org/en-US/docs/Web/JavaScript, Date: 2022-06-24.

[Co20] Commission, International Electrotechnical: Nuclear power plants – Instrumentation, control and electrical power systems – Security controls (IEC 63096). 2020.

[De] Design Disciplines, URL: https://element-plus.org/en-US/guide/design.html, Date: 2022-06-24.

[In] Introduction to Vue.js, URL: https://vuejs.org/guide/introduction.html, Date: 2022-06-24.

[IS17] ISO/IEC: Information technology - Security techniques - Application security, Part 5: Protocols and application security controls data structure (ISO/IEC 27034-5). 2017.

[IS18] ISO/IEC: Information technology — Application security,Part 5-1:Protocols and application security controls data structure, XML schemas (ISO/IEC 27034-5-1). 2018.

[Mo] MongoDB Manual, URL: https://www.mongodb.com/docs/manual/, Date: 2022-06-24.

[No] Node.js Documentation, URL: https://nodejs.org/api/documentation.html, Date: 2022-06-24.

[Ty] Typescript for the New Programmer, URL: https://www.typescriptlang.org/docs/handbook/typescript-from-scratch.html, Date: 2022-06-24.

[Wh] What is a cyber security posture and how do you assess it, URL: https://tinyurl.com/yww68vfp, Date: 2022-06-24.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1565

Towards Integrating Multi-Agent Organizations in OPC UA for Developing Adaptive Cyber-Physical Systems

Jan Sudeikat [1], Michael Kohler-Bußmeier[1]

Abstract: Developing adaptive Cyber-physical Systems (CPS) requires integrating physical components with run-time controls which monitor and adjust the operation of components appropriately. The use of multiagent systems (MAS) is an established approach for embedding adaptive features and the integration of multiagent systems in industrial systems is an active field of research. Industry standards, like OPC UA, allow for vendor independent information exchange and unit control in industrial settings. Here, we propose to represent MAS organizations within OPC UA information models. The organization of a MAS can be seen as a crosscutting concern in structing and integration distributed adaptive applications. Explicit modelling allows to express organizational structures in industrial settings and enables monitoring of these structures at run-time with established tools and frameworks for industrial control.

Keywords: Multiagent System, OPC UA, Organizational Model, Cyber-physical System.

1 Introduction

Multiagent systems (MAS) realize application functionality by the interplay of autonomous, interactive, rational agents. Utilizing agent technology and multiagent systems in industrial systems is an active field of research [Ka20]. These efforts particularly support designing Cyber-physical Systems (CPS), which integrate physical and digital components into coherent applications. Example application areas are i. a. Smart Grids and Industry 4.0 systems [Cr19]. Designing these systems requires integrating heterogeneous system entities and enabling adaptive features, e.g. robust and flexible operation of devices, which respond to local execution constraints and environmental contexts.

Open Platform Communications Unified Architecture (OPC UA)[2] is a standardized, cross-platform, international, and widely spread communication protocol for accessing sensor information and controlling devices in industrial settings (IEC 625413). A major design rationale is the adoption of a service-oriented approach for controlling industrial units. In order to facilitate interoperability, a standardized, graph-based modelling framework and extensible type-system enables the explicit provisioning of information models. These models describe the available information, e.g., sensor readings, set points, etc. [Go20, Iv21]. Open-source implementation frameworks are available [Mu21].

[1] Department of Computer Science, Hamburg University of Applied Sciences, Berliner Tor 7, 20099 Hamburg, Germany {jan.sudeikat|michael.koehler-bussmeier}@haw-hamburg.de
[2] https://opcfoundation.org/about/what-is-opc/

doi:10.18420/inf2022_135

In a prior work [SK22], we outlined the benefits of organizational models in (adaptive) CPS development. Interdisciplinary multilevel modeling efforts benefit form a crosscutting view that integrates systems in a coherent system of systems. We particularly aim at extending engineering approaches for CPS, e.g. following the *Use Case Methodology* [Go17]. While organizational concepts support coherent system designs, a fully-fledged development process also addresses integrating organizational concepts in the actual application [SK22]. Here, we outline work in progress on integrating organizational models in OPC UA information models. This enables to describe and modify logical organizations at run-time. In addition, these models can be accessed and inspected using established protocols and tools in industrial contexts.

This paper is structured as follows. In the next section we outline the role of organizational modelling in the context of building CPS. In Section 3, we outline work in progress on expressing these models within the OPC UA information models, before we conclude and give prospects for future work.

2 Organizational Modelling in CPS Development

In agent-based applications, the application functionalities arise from the (microscopic) interplay of agents. In *organization-centered multiagent systems* (OCMAS) explicit organizational models prescribe and constrain the interactions among agents, as well as their interactions with their environment / context. These models provide means for defining the macroscopic applications behaviour [DP13].

CPS-development benefits form integrating an organizational modelling perspective [SK22]. Coherence of the interplay of physical and digital system elements can be supported by establishing appropriate design abstractions. Group / Team abstractions are used to denote how agents can interact. A common concept is the *Role* agents play in a specific group. *Structural*, *functional*, and *deontic* dimension are used to describe how an organization operates [Hü10]. Thus, it can be regarded as a blueprint which allows to infer run-time aspects of the system. Using this abstraction, it is a design-challenge to balance coherence (global) and agent autonomy (local). Adaptability at run-time can be achieved by explicitly altering the organizational structure itself [KW13].

Our model definition is based on the fundamental Agent/Group/Role (AGR) model [Fe04]. Due to its generic structure, this model it can serve as a foundation for application dependent, elaborate models. In addition, usages to model cyber-physical systems have been reported, e.g. [Se17]. Figure 1 shows an AGR-based organization, where agents also integrate physical devices. The so-called *cheeseboard* notation from [Fe04] is used to denote the interplay of Agents, Groups and Roles. Agents may participate in several groups and can interact with each other. The notation is extended to denote the interactions of agents with physical devices. Agents can access physical devices and exert control (filled arrows in Figure 1) following integration modes as described in IEEE 2660.1 [Le21]. Here, single agents are responsible for controlling

physical units. E.g. the groups G1 and G2 may follow *Resource Agent* Pattern [Cr19]. Groups may also include the collaboration of agents for controlling units as described by the additional patterns in [Cr19]. In addition, members of the group G_3, playing the role *P* represent sets of physical assets. This may resemble the establishment of a peer-to-peer trading system for energy [Hu21], where agents playing the Role *P* trade energy and trading is supervised by a managing agent (Role *M*), e.g. as motivated in [SK22].

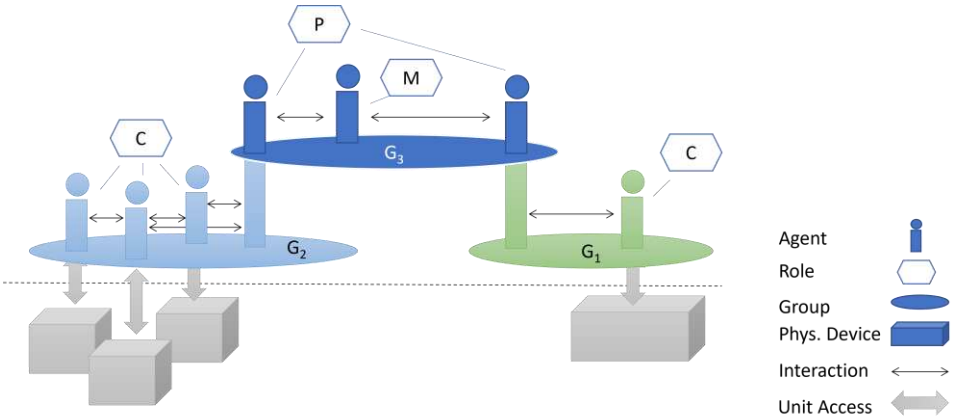


Fig. 1: AGR-based organization model structures the agent-based control of physical devices.

3 OPC UA Integration

OPC UA provides a comprehensive framework for exchanging information in industrial settings and controlling industrial devices. Servers provide data, in addition, methods can be invoked on these servers. A graph-based information model describes i. a. the available data, meta-data and executable functions (e.g. see [Go20, Iv21] for details).

We reviewed works integrating multiagent Systems with OPC UA. Despite the prominence of OPC UA in industrial control systems the integration of MAS has been addressed in selected works only. The seminal work [Ho17] showed that MAS can be integrated in OPC UA, including agent types and inter-agent communication. It has been shown that OPC UA can be used to decouple physical devices and MAS, particularly enabling MAS that are independent from underlying automation systems [Se21] and allows integrating differing frameworks [BV20]. Our approach builds on these efforts, since OPC-based information models are used to supplement high-level information about the organizational structure of the applications. To the authors knowledge, *explicit* modelling of mutable MAS organizations in OPC UA has not been addressed yet.

A generic structure for an organizational model, defined in terms of an OPC UA information model is given in Figure 2. Typically, organizational models are defined in

terms of object-oriented meta-models [Fe04]. Thus, our integration of these models in OPC UA builds on works which study how to translate object-oriented structures in the specifics of OPC information models [Le17,Pa18]. E.g. in [Pa18], it has been shown that compositions can be described by OPC UA folder structures. Thus, *Groups* aggregate *Roles* and *Interactions*. Additional functions and information, e.g. a *MemberCount*, can be added as sub-nodes as needed. Sub-elements of organizational Roles denote which agents play a specific Role. *RoleConstraints* indicate structural constraints on roles, e.g. the *dependence* of additional memberships or the *correspondence* of Roles [Fe04]. *Interactions* denote communication patterns among Roles.

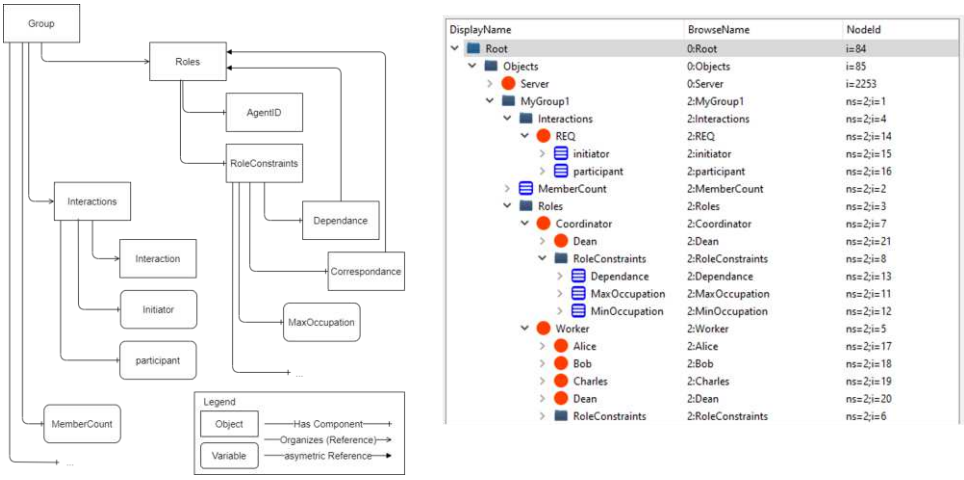


Fig. 2: Left: Information model excerpt, using OPC UA standard notation [IEC21]; Right: Demonstrative instantiation of a corresponding organizational model[3]

Providing organizational models in an OPC-based application setting requires deployment in an OPC UA server. Three major options for this provisioning are (1) hosting the server within a dedicated agent, (2) hosting a dedicated server in the environment or (3) hosting the model as a supplemental model within already existing servers. Deployment is an architectural choice, independent of the contents of the model. Due to the availability of lightweight server implementations in various programming languages [Mu21], the second option is used in our proof of concept. This corresponds to the Agent & Artifacts approach, which allows to provide organizational models [Hü10].

A major motivation for explicit modelling of organizational structures it to enable *transparency* and *adaptability*. Organizations can be understood as *static* structures, which describe agent collective, e.g. using established patterns [Sa19]. This corresponds to a permissioned model on servers, which is set up at design time, can be inspected using OPC UA clients, and would only be altered by privileged agents or users. A major

[3] Displayed using the freely available client from: https://github.com/FreeOpcUa/opcua-client-gui

requirement for *adaptive* systems would be to adjust organizations at-runtime, as agents enter and leave the system or internally adjust their execution. This can be achieved by adding / removing the corresponding nodes, e.g. indicating group membership. Based on that, Groups and Roles can be created interactively, e.g. by direct interaction with OPC UA servers and creating organizational structures on the fly. Following this approach, agents can maintain and manage information nodes themselves. Future work comprises enabling open systems, where these modifications would have to be constrained and validated, e.g., using OPC UA-based authentication and/or guarding modifications.

4 Conclusions

In this paper, we have argued that OPC UA provides suitable abstractions to describe organizational models in industrial systems. While prior works have studied how agents can be embedded in CPS using OPC UA, we propose here to supplement organizations as a logical view on inter-agent structures. Future work will address the pragmatics of integrating OCMAS in OPC UA. While freely available frameworks [Mu21] allow programming level integration in agent-based systems agent development would benefit from a suitable programming model for describing and altering organizations at run-time. In addition, the integration of organizational modelling in CPS development practices, e.g. based on the *Use Case Methodology* [Go17], is to be explored.

Bibliography

[BV20] Baumgartel, H., Verbeet, R.: Service and Agent based System Architectures for Industrie 4.0 Systems. In: NOMS 2020 - 2020 IEEE/IFIP Network Operations and Management Symposium, IEEE, Budapest, Hungary, pp. 1–6. 2020.

[Cr19] Cruz Salazar, L.A., Ryashentseva, D., Lüder, A., Vogel-Heuser, B.: Cyber-physical production systems architecture based on multi-agent's design pattern - comparison of selected approaches mapping four agent patterns, Int J Adv Manuf. Technol 105, 2019.

[DP13] Dignum, V., Padget, J.: Multiagent Organizations. In: Multiagent Systems. Second Edition, MIT Press, 2013

[Fe04] Ferber, J., Gutknecht, O., Michel, F.: From Agents to Organizations: An Organizational View of Multi-agent Systems. In: Giorgini, P., Müller, J.P., Odell, J. (Eds.), Agent-Oriented Software Engineering IV, Springer, pp. 214–230, 2004.

[Go17] Gottschalk, M.; Uslar, M. & Delfs, C.: The Use Case and Smart Grid Architecture Model Approach, *Springer International Publishing,* 2017

[Go20] Gong, Y., Wang, Z., Han, D.: OPC UA Information Modeling Method and Xml Definition. In: 2020 IEEE Conf. on Telecommunications, Optics and Computer Science (TOCS). IEEE, Shenyang, China, pp. 328–331, 2020.

[Ho17] Hoffmann, M.: Adaptive and Scalable Information Modeling to Enable Autonomous

Decision Making for Real-Time Interoperable Factories. PhD Thesis, RWTH Aachen University, DOI: 10.18154/RWTH-2017-07373, 2017.

[Hu21] Huang, Q., Amin, W., Umer, K., Gooi, H.B., Eddy, F.Y.S., Afzal, M., Shahzadi, M., Khan, A.A., Ahmad, S.A.: A review of transactive energy systems: Concept and implementation. Energy Reports, https://doi.org/10.1016/j.egyr.2021.05.037, 2021.

[Hü10] Hübner, J.F., Boissier, O., Kitio, R., Ricci, A.: Instrumenting multi-agent organisations with organisational artifacts and agents: "Giving the organisational power back to the agents." Auton Agent Multi-Agent Syst 20, 369–400, 2010.

[IEC21] IEC 625413:2020, OPC Unified Architecture – Part 3: Address Space Model, 2021.

[Iv21] Ivanova, T., Batchkova, I., Gocheva, D., Belev, Y.: Information Modeling of Cyber-Physical Systems using OPC-UA. In: 2021 International Conference Automatics and Informatics (ICAI). IEEE, Varna, Bulgaria, pp. 43–47, 2021.

[Ka20] Karnouskos, S., Leitao, P., Ribeiro, L., Colombo, A.W.: Industrial Agents as a Key Enabler for Realizing Industrial Cyber-Physical Systems: Multiagent Systems Entering Industry 4.0. IEEE Ind. Electron. Mag. 14, 18–32, 2020.

[KW05] Köhler-Bußmeier, M.; Wester-Ebbinghaus, M.: Model-driven middleware support for team-oriented process management. Transactions on Petri Nets and Other Models of Concurrency, 8:159–179, 2013.

[Le17] Lee, B., Kim, D.-K., Yang, H., Oh, S.: Model transformation between OPC UA and UML. Computer Standards & Interfaces 50, 236–250, 2017.

[Le21] Leitão, P.; Strasser, T.; Karnouskos, S.; Ribeiro, L.; Barbosa, J. & Huang:, V.: Recommendation of Best Practices for Industrial Agent Systems based on the IEEE 2660.1 Standard. Proc. of the 22nd IEEE Int. Conf. on Ind. Techn. (ICIT), 2021

[Mu21] Mühlbauer, N.; Kirdan, E.; Pahl, M.-O. & Waedt, K.: Feature-based Comparison of Open Source OPC-UA Implementations, INFORMATIK 2020, GI, 367-377, 2021.

[Pa18] Pauker, F.; Wolny, S.; Fallah, S.M.; Wimmer, M.: UML2OPC-UATransforming UML Class Diagrams to OPC UA Information Models. Procedia CIRP 67, 128–133, 2018.

[Sa19] Cruz Salazar L.A.; Ryashentseva, D.; Lüder, A.; Vogel-Heuser, B.: Cyber-physical production systems architecture based on multi-agent's design pattern—comparison of selected approaches mapping four agent patterns. Int J Adv Manuf Technol 105, 2019.

[Se21] Seitz, M., Gehlhoff, F., Cruz Salazar, L.A., Fay, A., Vogel-Heuser, B.: Automation platform independent multi-agent system for robust networks of production resources in industry 4.0. J Intell Manuf 32, 2023–2041, 2021.

[Se17] Seddari, N.; Belaoued, M.; Bougueroua, S.: Agent/Group/Role Organizational Model to Simulate an Industrial Control System. In: Int. Journal of Social, Behavioral, Educational, Economic, Business and Ind. Eng., vol. 11, no10, pp.2376–2385. 2017.

[SK22] Sudeikat, J.; Köhler-Bußmeier, M.: On Combining Domain Modeling and Organizational Modeling for Developing Adaptive Cyber-Physical Systems. In: *Proceedings of the 14th International Conference on Agents and Artificial Intelligence - Volume 1: ICAART,* ISBN 978-989-758-547-0, pages 330-336, 2022.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1571

Interactive graphical modeling of security artefacts for abstracted Industry 4.0 automation systems

Louis Roger Tchuegoue Djeukoua[1] Edin Kreho[2] Siwar Belaidi[3] Karl Waedt[4]

Abstract: The frontend and backend are found in all software and therefore also on all websites. These two terms describe two different layers that make up programs or pages. About two thirds of all companies have their own website, and most employees use computers.Globally, cyberattacks are becoming more prominent and spreading to multiple areas, and the move to Industry 4.0 requires increased security measures. Important security precautions must be taken from the development of industrial devices that use the Industrial Internet of Things, with IEC 62443, ISO/IEC 27001, and their integration into the architecture of existing information and automation systems must be secure. IEC 62443 focuses on the IT security of so-called industrial automation and control systems (IACS), which are necessary for the safe and reliable operation of automated factories or infrastructures. [Her08] Since security breaches are inevitable, it is also important to implement detection and response mechanisms in industrial automation and control systems (IACS). Together, these measures will enable various organizations to achieve an appropriate level of resilience. This paper discusses the interactive graphical representation of large-scale industrial automation systems for the purpose of modeling and evaluating cybersecurity during all phases of the industrial equipment life cycle. In addition, it addresses the expressiveness and scalability of front-end graphical problems by assuming that a multi-user back-end server with a semi-formal representation of cybersecurity-related artifacts is available, at least in software prototype form.

Keywords: Industrial Automation and Control System; Web Framework; Graphical User Interface; bi-directional data binding; cooperative real-time development; Cybersecurity; Attack Scenarios; Cyber Physical System

1 Introduction

Advances in the Industrial Internet of Things (IIoT), interconnectivity, continued growth of smart ecosystems, and quintessential process automation have not only opened up immense opportunities, but also expanded the cyber threat and attack landscape. The main

[1] Hochschule für Technik und Wirtschaft Berlin (HTW Berlin), Fachbereich 1 - Energie und Information, Treskowallee 8, 10318 Berlin, Germany, rogerdjeukoua@yahoo.fr
[2] Friedrich-Alexander-Universität Erlangen(FAU Erlangen), Lehrstuhl Informatik Fau Erlangen, Schlossplatz 4, Erlangen, Germany edin.kreho@fau.de
[3] Framatome GmbH,ICETA-G Department ,Paul-Gossen-Straße 100,91052, Erlangen,Germany siwar.belaidi@framatome.com
[4] Framatome GmbH, ICETA-G Department, Paul-Gossen-Straße 100, 91052, Erlangen, Germany karl.waedt@framatome.com

doi:10.18420/inf2022_136

manifestation of this shift is that industries are striving to align with Industry 4.0 goals on demand. In particular, it has become clear that the convergence and amalgamation of industrial operational technology (IoT) and information technology (IT) is driving the sophistication, complexity of the IIoT ecosystem, paradigm shift, and global security changes.The more innovative companies become, the greater their vulnerability to attack. The failure of a single factory can result in losses of millions of dollars or Euros per day or even endanger human lives. Therefore, it is critical to minimize the risks of cyber attacks. [Her08] [Ple19]

2 Problem Statement

Industry 4.0 and other IoT environments are typically composed of many different components such as sensors, pumps, pressurizers, valves, motors, circuit breakers and other similar devices. These frequently exchange data, building up a complex and confusing information network. In the event of an intrusion, it is important to quickly assess which data could be compromised. However, due to insufficient documentation and analysis techniques, companies and institutions often have difficulties or it takes a very long time to find out which data is actually affected.Based on the modeling, analysis tools can then be used to quickly assess how attacks on abstracted industrial automation systems can be avoided or eliminated. This paper addresses the interactive graphical representation of large industrial automation systems with the goal of cybersecurity modeling and assessment during all lifecycle phases of industrial equipment. The graphical representation includes primary and supporting assets as defined in ISO/IEC 27005:2022. [22] (See fig. 1)

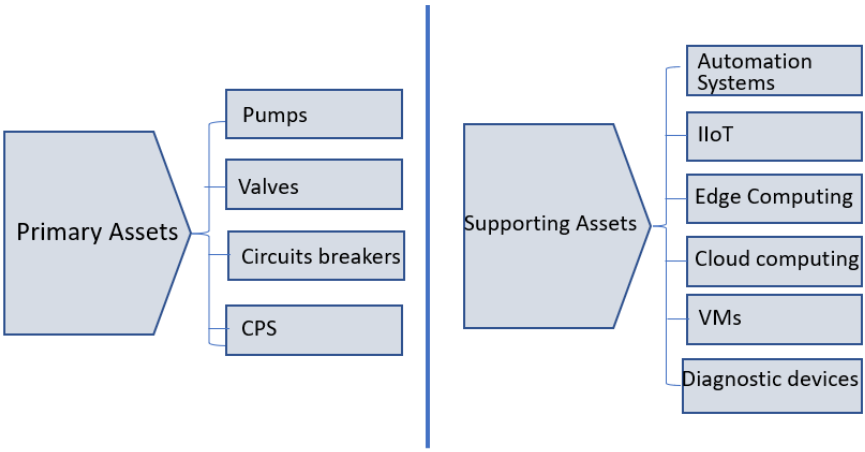


Figure 1: Primary and supporting assets mentioned in this paper

As described on the figure above, primary assets addressed and graphically modeled in this paper include pumps, valves, circuit breakers, and general Cyber Physical Systems (CPS) that interact with automation systems. Supporting assets are distributed automation systems, industrial IoT systems (IIoT), edge computing, cloud computing, virtual machines, and diagnostic devices that are only temporarily connected to other supporting assets.[Wik22][Beh17] In this paper, only the expressiveness and scalability issues of the graphical front-end are addressed. It is assumed that a multi-user capable backend server with a semi-formal representation of cybersecurity-related artifacts is available, at least as a software prototype[Kie13].

3 Abstracted Industry 4.0 automation systems

3.1 Definition

According to DIN 19 233, the term automation is understood to mean the “equipment of a device so that it operates wholly or partly as intended without the assistance of a human being”. It is also defined as the use of artificial systems that automatically follow a program and make decisions on the basis of the program to control and, if necessary, regulate processes [Vet82].

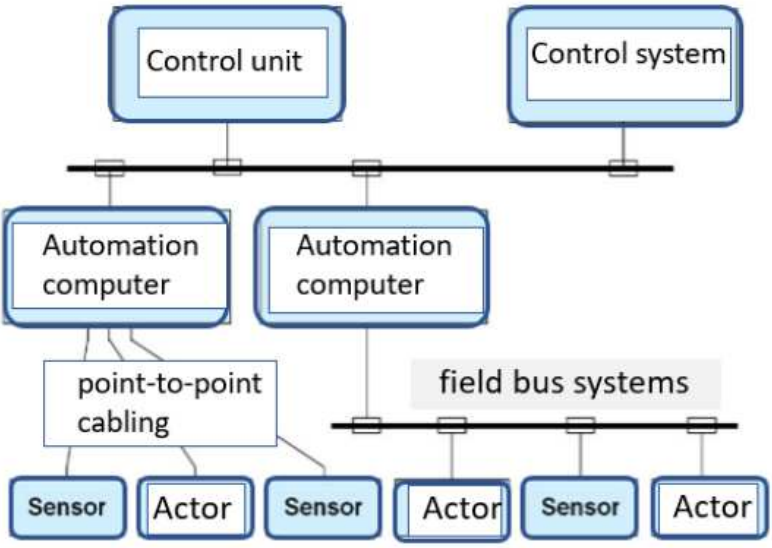


Figure 2: Basic structure of example automation systems

Additionally, the term automation includes the assumption of process control and possibly process regulation tasks by artificial systems. [Her08] The figure 2 above describes the basic structure of an example automation systems. An Abstracted Industrial Automation System is a conceptual generalized automation platform that allows the representation of different concrete industrial automation systems. The concept of industrial automation system in this paper is not about an automation system from a specific manufacturer but a fully configurable or meta automation system that is not tied to a specific manufacturer.[Wit22] In the sense of Cloud Computing Reference Architecture or Big Data Reference Architecture or IoT Reference Architecture. The concept approaches Automation Architecture what means that the automation system will assume certain characteristics, such as:

- an automation platform that allows multiple automation systems to be configured
- the automation platform supports the implementation of networked automation systems
- each automation system consists of automation subsystems
- each subsystem contains several hardware modules (assemblies or compact hardware units) The hardware modules are also called embedded components
- A hardware module optionally belongs to a subrack (rack/subrack)
- The hardware executes real-time software
- A hardware subsystem is optionally located in a IC cabinet (which is optionally lockable and optionally indicates whether the cabinet door is open.
- Optionally, a DCS cabinet can have two doors (front door to plug in the boards and rear door to perform wiring)
- The embedded subsystems are optionally connected to service devices or diagnostic equipment (usually notebooks). Service devices can remain continuously connected. Diagnostic devices are only connected temporarily, e.g. for initial software loading, parameterization and troubleshooting.
- The control cabinets (or directly the embedded systems) are located in a lockable (or not lockable) room.
- The rooms are located on a certain corridor (level) of a building.
- A facility (power plant or factory) may contain multiple buildings. (Separate buildings or rooms are often required for safety reasons, e.g., so that not all buildings are affected in an assumed unlikely airplane crash.
- All hardware components have certain interfaces that are relevant for security considerations, e.g. RJ45 for network communication via copper cables, optical connections, USB interfaces, serial interfaces (especially embedded systems), ...

- Security measures (security controls) are applied at all levels (building, hallway, room, control cabinet, ... , application software).

3.2 Application fields of Abstracted Industry 4.0 automation systems

Industry 4.0 includes cyber-physical systems (CPS), the Internet of Things (IoT), the Industrial Internet of Things (IIOT), cloud computing, cognitive computing and artificial intelligence.In addition to purely industrial applications, such as intelligent manufacturing and production environments in various industries (smart factories), cyber physical systems are also used in many other areas[Vet82]. These include intelligent power grids (Smart Grids), electronic health, age-appropriate assistance systems, but also intelligent traffic monitoring systems or automatic early warning systems in disaster control.Industrie 4.0 and CPS exhibit a high degree of networking and operate largely autonomously. One of the greatest challenges is to establish standards and industry grade products that ensure the interoperability of cyber-physical systems. Only seamless interaction between the various technologies and hardware or software components will enable Industrie 4.0 to realize its full potential. Cross-company networking of cyber-physical systems is necessary to control complex plants and processes.[Pre16]

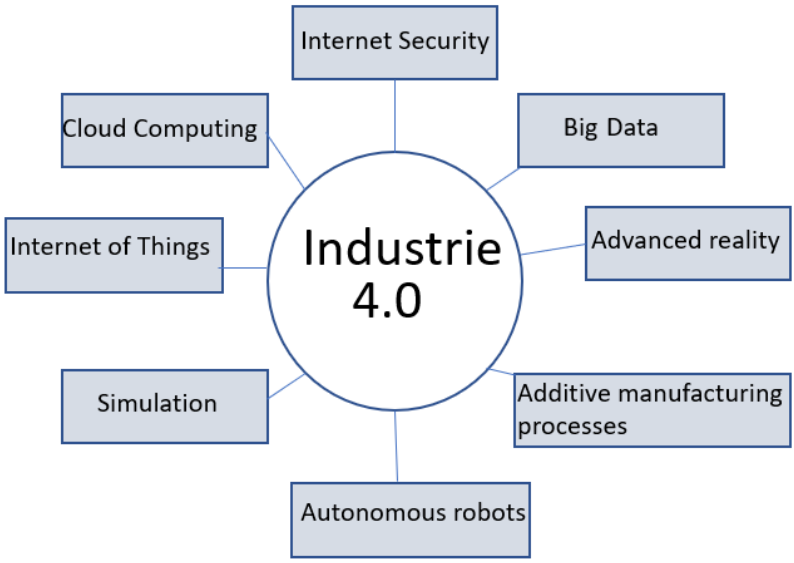


Figure 3: Application fields of industry 4.0

3.3 Interconnection and structuring of UI model with interactive assets representation

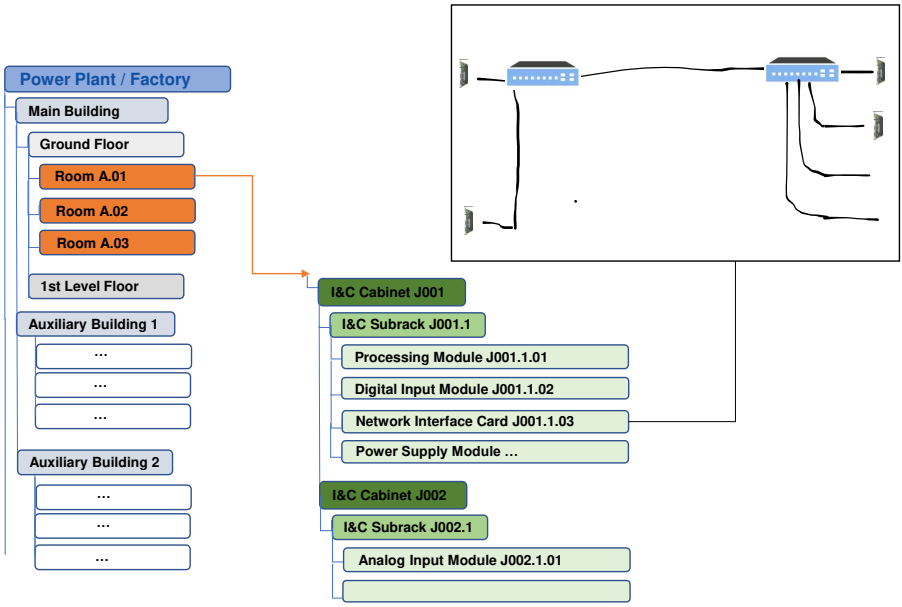


Figure 4: Assure UI top-level – tree view example

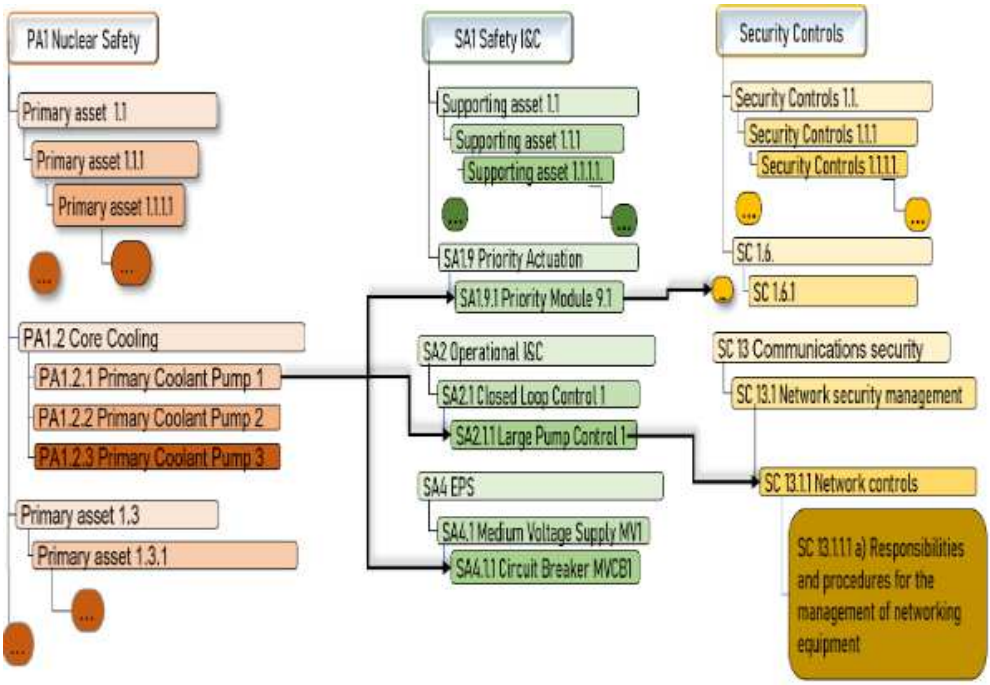


Figure 5: Interaction among primary assets, supporting assets and security control

The diagram in Fig. 5 illustrates all assets and their relationships.Several types of security controls are available, such as: categories - objectives - security controls - security sub-controls.For example: Communications Security - Network Security Management - Network Controls - Responsibilities and Procedures for Managing Network Equipment). However, assets are exposed to potential attacks, so it is critical to structure the primary and supporting assets so that a target security level can be assigned to the primary assets so that security controls are properly modeled and implemented.

4 Web Framework

4.1 Definition within the intended industrial application scope

A framework is a structure that you can build software on. Vue.js 3 proved once again how simple and fast JavaScript frameworks can and should be. Whether professional app development or in the private sector - Vue.js has many industry proponents and is becoming increasingly industry mainstream [Jos21]. With Vue.js version 3, many new features and optimizations appeared that raise the JavaScript framework to a new level. No

wonder many consider Vue.js one of the easiest and most accessible frameworks. For example, Vue allows direct support for HTML templates, while tools like React still have to define Digital Optics Monitoring (DOM) elements and use JavaScript for this. It provides various tools to help the developer in his work. This is not only about the different graphical elements, but also about what happens in the background.One of the main advantages of Vue.js is its ability to provide very simple responsive two-way data binding for custom form inputs and Vue components. There are also special libraries (that can be used in Vue.js) that provide pre-built components, graphical elements (controls) (Matthias Jost, 2021). Because Vue 3 is smaller and faster than Vue 2, it provides some low impact APIs and offers improved TypeScript support and a more maintainable code base[Sto17].

4.2 Graphical User interface of the considered Web Framework

A graphical user interface (GUI) is an interface that allows the user to interact with various electronic devices and industrial processes using icons and other visual indicators. The graphical user interfaces were created because the command line interfaces were quite complicated and it was difficult to learn all the commands they contained. Also, the visual feedback and structuring are essential for scalability. Within this paper, the library Elementplus will be used to create to design the GUI of the considered Web Framework. Element Plus is a component library offering a cool design language, many customization options and detailed documentation. It builds on Vue 3.0 for developers, designers and product managers and uses TypeScript, the composition API to create a better experience for developers.[Plu]

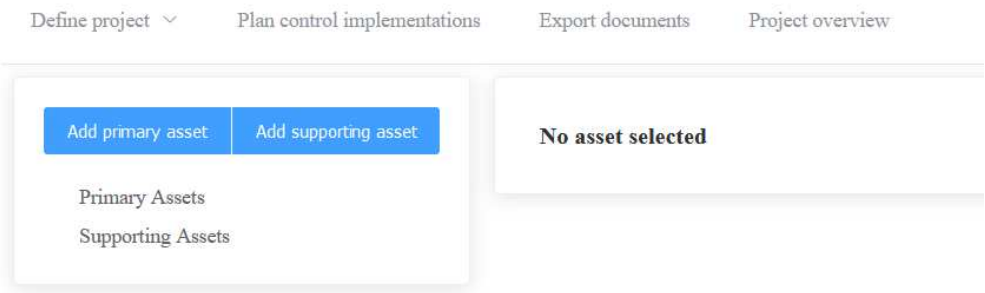


Figure 6: Initial Minimalist Graphical User Interface

5 Bi-directional Data Binding

When a bidirectional data binding is requested, the runtime environment retrieves the source value and initializes the specified target property with that value. Each time the source value changes, the data binding retrieves the new value and reinitializes the target property. When

the value of the target property changes, for example, when the user makes an input to an edit control, the data binding retrieves the new value of the target property and transfers it back to the source. The two-way data binding is the default type of data binding[Mic22]. The bi-directional data binding is especially important when multiple clients (as indicated in Fig. 7) are simultaneously modelling security artefacts of a plant or factory.

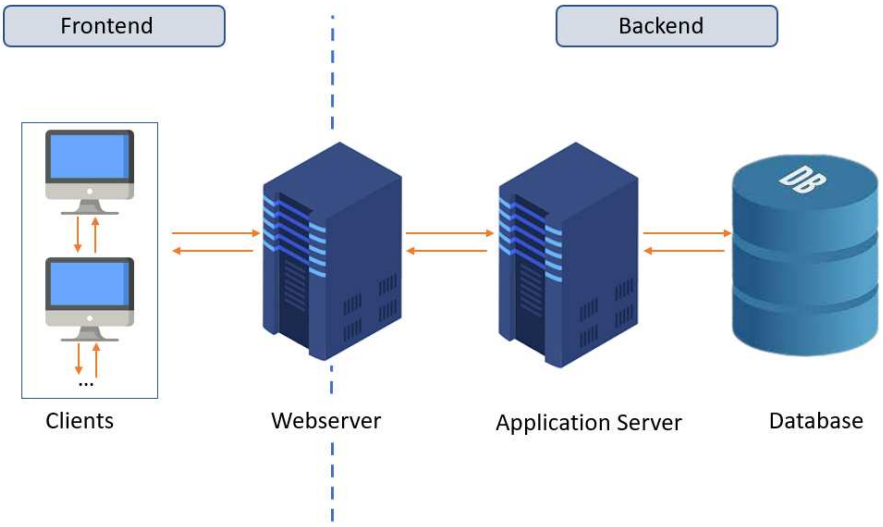


Figure 7: Bi-directional Data Binding

5.1 Backend data management

MongoDB is an open source, non-relational database management system (DBMS) that uses flexible documents instead of tables and rows to process and store various types of data. It is a schema-free document database written in C++, which is being developed in an open source project primarily run by 10gen Inc, that also provides professional services related to MongoDB. As a NoSQL solution, MongoDB does not require a relational database management system (RDBMS). Therefore, it provides an elastic data storage model that allows multivariate data types to be easily stored and queried by users. This not only simplifies database management for developers, but also creates a highly scalable environment for cross-platform applications and services. The main goal of MongoDB,

according to its developers , is to bridge the gap between fast and highly scalable key/value stores and feature-rich traditional RDBMSs (Relational Database Management Systems). MongoDB databases are located on a MongoDB server, which can host several such databases that are independent of each other and stored separately by the MongoDB server. Typically, a database contains one or more collections consisting of documents. A number of database security credentials can be defined to control access to the database. [DB]

5.2 The IoT Attack Lifecycle

The attack lifecycle describes the process by which attackers can typically gain access to organizations, their networks, systems and data.



Figure 8: Stages of the IoT attack lifecycle [Pal22]

6 Scalability of Graphical Front-end

Scalability generally refers to the ability of an organization, network, or process to evolve and be able to keep up with increased demand. Scalability is one of the most critical requirements for the success of a web application. When a company's front-end becomes scalable, it automatically gains an advantage over its competitors because it can now easily adapt to the changing needs of customers and/or consumers. This not only shows that the company is competitive, but also that it is stable and ready to deal with a sudden surge in demand, changing needs and trends, and the appearance of new competitors in the market.

7 Conclusion

The progress of digitalization is accompanied by various solutions in the field of automation. Cybersecurity plays a first-class role not only in Industry 4.0, but also in all IT areas because without cybersecurity, digitization will not be successful. Security measures aim to protect the confidentiality, integrity and availability of an IT system from being compromised by deliberate or accidental attacks. process and security improvement Security improvement must be a continuous activity. Cybersecurity must be a priority for Industry 4.0 organizations, from manufacturers to critical national infrastructure. The structured, scalable and interactive representation of security artefacts for simultaneous use by multiple cybersecurity team members is intended to cater towards more comprehensive and more effective planning, implementation and tracking of cybersecurity related artefacts.

Literatur

[22] *ISO/IEC 27005:2022 -Information security, cybersecurity and privacy protection — Guidance on managing information security risks*. en. Standard. International Organization for Standardization, 2022. URL: https://www.iso.org/standard/80585.html.

[Beh17] Jan Odenbach Dr. Edgar Göll Dr. Siegfried Behrendt. „Industrie 4.0 – digital-vernetzte dezentrale Produktion“. In: *https://www.researchgate.net/* 47 (Jan. 2017).

[DB] MONGO DB. *Build faster Build smarter*. URL: https://www.mongodb.com/docs/manual/introduction/.

[Her08] Christof Hübner Hermann Merz Thomas Hansemann. *Building Automation*. 2008. ISBN: 978-3-410-16760-0. DOI: https://10.1007/978-3-319-73223-7. URL: https://link.springer.com/book/10.1007/978-3-319-73223-7.

[Jos21] Matthias Jost. *Enthusiastic About Software Development*. 2021. URL: https://www.matthias-jost.ch/warum-vuejs-einfach-erlernbar-ist/#:~:text=Vue.js%20dient%20dazu%20grafische%20Oberfl%C3%A4chen%20%28Frontends%29%20f%C3%BCr%20Webapplikationen,grafischen%20Elemente%2C%20sondern%20darum%2C%20was%20im%20Hintergrund%20passiert..

[Kie13] Jan U. Kieß. *Objektorientierte Modellierung von Automatisierungssystemen*. 2013. URL: https://link.springer.com/book/10.1007/978-3-642-79905-1.

[Mic22] Microsoft. *Data Binding*. 2022. URL: https://docs.microsoft.com/en-us/dotnet/desktop/wpf/data/?view=netdesktop-6.0.

[Pal22] Palaolto. *Impacts of Cyberattacks on IoT Devices*. 2022. URL: https://www.paloaltonetworks.com/apps/pan/public/downloadResource?pagePath=/content/pan/en_US/resources/research/impacts-of-cyberattacks-on-iot-devices.

[Ple19] Valentin Plenk. *Grundlagen der Automatisierungstechnik kompakt*. 2019. ISBN: 978-3-658-24469-9. URL: https://link.springer.com/book/10.1007/978-3-658-24469-9.

[Plu] Element Plus. *Element Plus - a Vue 3 based component library for designers and developers*. URL: https://element-plus.org/en-US/.

[Pre16] Giovanni Prestifilippo. *Auswirkungen von Industrie 4.0 auf Warehouse-, Transport- und Supply-Chain-Management-Systeme*. Springer, 2016. ISBN: 978-3-662-53250-8.

[Sto17] Christopher Stock. *A Review Paper on Botnet and Botnet Detection Techniques in Cloud Computing*. 2017. URL: https://blog.mayflower.de/6135-typescript.html.

[Vet82] Günter Vettin. In: *Verfahren zur technischen Investitionsplanung automatisierter flexibler Fertigungsanlagen*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1982, S. 130–138. ISBN: 978-3-642-81904-9. DOI: 10.1007/978-3-642-81904-9_11. URL: https://doi.org/10.1007/978-3-642-81904-9_11.

[Wik22] Wikipedia. *Industrial internet of things*. Juni 2022. URL: https://en.wikipedia.org/wiki/Industrial_internet_of_things.

[Wit22] Christoph Witte. *A Review Paper on Botnet and Botnet Detection Techniques in Cloud Computing*. 2022. URL: https://it-rebellen.de/2019/01/16/bedenkliche-einstellung-zur-cybersecurity-von-industrieanlagen-risiken-werden-unterschaetzt/.

New cybersecurity standards for IACS of the nuclear power industry in China

Yun Guo[1], Junjie Wang[1]

Abstract: As of June 2022, China has ranked second with respect to the number of nuclear power plants (NPPs) under construction and in operation in the world. It's expected that China's installed nuclear power capacity will reach about 70 million kilowatts by 2025 [1]. At present, industrial control systems of China's new NPPs have adopted digital industrial ones, and existing NPPs are gradually evolving from a mixture of digital and analog to full digital technology, which has brought great challenges to the cyber security of nuclear power plants. The cyber attack on the industrial control systems of NPPs may not only lead to interruptions in the production process, but also may cause nuclear safety incidents. Therefore the Standardization Administration of the People's Republic of China and relevant industry regulatory departments have respectively developed national and industry level standards regarding the cybersecurity of industrial control systems of nuclear power plants in recent years. This report introduces China's new standards on the cyber security of industrial control systems in NPPs, analyzes their relationship with relevant international standards, puts forward the issues that need to be considered regarding the coordination of those standards, and provides reference for the subsequent development of new international standards.

Keywords: nuclear power plants, cybersecurity standards, nuclear safety, Industrial Automation and Control System, I&C

1 Introduction

The Industrial Automation and Control Systems (IACSs) of the nuclear power plants (NPPs) perform the monitoring, control and protection functions of the entire NPP. For a long time, IACSs of NPPs have physically isolated from other external systems and use dedicated hardware and software to run proprietary protocols, so it is viewed traditionally that the IACSs of NPPs are less likely to be affected by cyberattacks. However, with the trend of digitalization, networking and intelligence of global industry, IACSs of NPPs have been facing more and more cyber challenges, mainly reflected in: First, with the need for digital development, the communication between IACSs and external systems has increased, such as the connection between a SIS and distributed control system (DCS) to obtain relevant production data. Secondly, there are also more ways to break through physical isolation, such as reserving backdoors and implanting viruses during the development phases, spreading worms through mobile media during the operational phase, stealing information through laptop access during the maintenance phase, and even breaking through physical protection systems through social engineering

[1] Huaneng Shandong Shidao Bay Nuclear Power Company, ShanDong, China

doi:10.18420/inf2022_137

methods. Finally, the proprietary nature of IACSs and protocols can only temporarily increase the difficulty of the attack and the threshold for the attack is constantly lowered, since IACSs begin to use more and more common schemes, such as industrial communication based on Ethernet and TCP/IP protocol, workstations based on Windows operating system, HMI based on web browser mode. Along with the improvement of the open interconnection, those general solutions also leads to the increasingly serious cybersecurity threats. From the "Stuxnet" incident in Iran in 2010 [2]to the nationwide blackout in Ukraine in 2016, events in recent years have shown that cybersecurity has become an important influencing factor in the operational safety as well as more serious nuclear safety consequences of nuclear facilities[3]. Therefore, the cybersecurity of IACSs of NPPs have attracted great attention of the nuclear energy related community.

To dress those issues, countries around the world and relevant international organizations have carried out relevant research on the issue of IACS cybersecurity in NPPs and formulated corresponding standards and regulatory requirements. For example, the International Electrotechnical Commission (IEC) began to develop an international standard IEC 63096 [4]in 2016, with the aim of providing detailed security control guidelines based on hierarchical and lifecycle-specific method for IACSs of NPPs, helping all stakeholders of NPPs prevent, detect and react to cyberattacks to ensure the availability, integrity and confidentiality of IACSs. Also, the U.S. Nuclear Regulatory Commission (NRC) has developed computer security guideline RG 5.71 [5]for nuclear facilities in accordance with the protection requirements of the Federal Regulations for digital computers, communication systems and networks, with the purpose of ensuring that digital computers and communication systems and networks related to nuclear facilities are adequately protected from cyber attacks. As the second country in the world in terms of the number of nuclear power units currently in operation and under construction, China has also made exploration of relevant standards in recent years. With regard to actual cybersecurity implementation, NPPs often need to consider multiple standards, thus making the coordination of different standards essential and critical. To this end, this paper introduces China's relevant standards on the cybersecurity of IACSs in the nuclear power industry, analyzes their relationship with relevant international standards. Based on that, the paper puts forward the problems that need to be considered when coordinating the use of those standards, providing a reference for implementing the security measures in NPPs as well as a reference for the subsequent development of relevant international standards.

2 New cybersecurity standards for IACSs of NPPs in China

2.1 Series of Standards for Classified Protection of Cybersecurity

Beginning in 1994 with the State Council's Decree “Regulations on the Security Protection of Computer Information Systems of the People's Republic of China”, China’s classified protection of cybersecurity was gradually promoted in an orderly

manner and was fully applied in various industries. The series of classified protection standards including more than 10 standards such as GB/T22239 [5]、GB/T22240 [6]、GB/T25070 [7]、GB/T 28448 [8]、GB/T 28449[9], which covers classification guide, baseline requirements, implementation guide, assessment requirements, etc. At the same time, the series of those standards have also been updated in a timely manner according to the new situation and new risks in cyberspace. Those standards and release time is listed in table 1.

Table 1: series of standards for classified protection of cybersecurity

Number	Name	Release time
GB/T 22239	Baseline for classified protection of cybersecurity	2019
GB/T 25070	Technical requirements of security design for classified protection of cybersecurity	2019
GB/T 28448	Evaluation requirements for classified protection of cybersecurity	2019
GB/T 28449	Testing and evaluation process guide for classified protection of cybersecurity	2018
GB/T 36958	Technical requirements of security management center for classified protection of cybersecurity	2018
GB/T 25058	Implementation guide for classified protection of cybersecurity	2019
GB/T 22240	classification guide for classified protection of cybersecurity	2020

The relevant standards for cybersecurity classified protection 2.0 have been released since May 2019, and a significant change in the new standards is to change the original security requirements to security general requirements and security extension requirements, of which the security general requirements are still divided into technical requirements and management requirements, but the lower-level requirements are changed from the previous standards. Under this framework, the technical requirements and management requirements both include five second-level requirements and they are listed as table 2.

Table 2: baseline requirements of for classified protection of cybersecurity

First-level requirements	second-level requirements
general technical	secure physical environment

requirements	secure communication network
	secure area boundary
	secure computing environment
	security management center
general management requirements	management system
	management organization
	management personnel
	construction management
	operation and maintenance management
extension requirements	cloud computing security
	mobile connecting security
	IoT security
	IACS security

The main differences between the new series and the previous series include:

- The new standards expand the scope incorporating cloud computing, mobile internet, Internet of Things, IACS etc. into the scope of standards.
- It is now recommended to use the triple protection architecture supported by secure communication network, secure boundary and secure computing environment along with the management center.
- The requirements of TC (Trusted Computing) technology are strengthened in this version.

2.2 GB / T 41241-2022 Management requirements for cybersecurity of industrial control systems in nuclear power plant

In 2022, the Standardization Administration of China issued GB/T 41241[10] "management requirements for cybersecurity of industrial control systems in nuclear power plant", which is based on the basic requirements of GB/T 22239 and combines the actual needs of NPPs to stipulate the requirements for the management, technical protection and emergency management for cybersecurity of industrial control systems NPPs. This standard applies to cybersecurity activities of all phases of the life cycle of IACSs of nuclear power plants (including design, development, engineering

implementation, operation and maintenance, decommissioning, etc.). It also applies to system maintenance activities that guide users of IACSs of NPPs to improve and enhance the cybersecurity protection capabilities.

This standard proposes the concept of cybersecurity fortification degree which is divided into 1-5 levels, and recommends that the fortification degree be classified according to the maximum consequences of the impact on nuclear safety and power generation after a successful cyber attack on the system. This standard provides a recommended correspondence between the cybersecurity fortification degree and that of the national classified protection series of standards.

3 Issues to be considered in the coordination of standards

3.1 Coordination of Security Degree

The most significant feature of IACSs in NPPs compared with traditional IT systems and IACSs of other industries is that the character of a hierarchical strategy used by IACSs in NPPs. To be specifically, IACSs of NPPs perform functions related to nuclear safety. From the design stage, IACSs of NPPs needs to implement the "classification strategy", that is, the instrumentation and control (I&C) function is first classified according to its nuclear safety importance, and then IACSs that perform different functional categories is classified according to the functional classification. Therefore, the cybersecurity of IACSs in NPPs need to consider its correlation with nuclear safety in terms of determining the degree of protection and the choice of protective measures.

At present, whether it is China's national standards or the relevant standards of international organizations such as IEC, they have put forward proposals for hierarchical protection for IACSs of NPPs, but the criteria for grading seems different intuitively. For example, IEC 63096 inherits the method of cybersecurity grading in IEC 62645[11]. Upon that, based on the classification of I&C functions and the analysis of the maximum consequences imposed on the nuclear safety level and performance of power plants after the successful implementation of cyberattacks, the cybersecurity degree of IACSs in NPPs is divided into S1, S2, S3. Further, IEC 63096 also defines a "BR" degree which means all IACSs in NPPs that are not classified as S1, S2, S3 are classified as BR and should follow the security control recommendations of the BR degree. In this sense, it can be considered that IEC 63096 includes 4 cybersecurity degrees for IACSs, whereas there is 5 degrees in China’s classified protection system and the cybersecurity fortification degree in GB/T 41241 is also divided into 5 levels. Therefore, how to establish an appropriate correspondence between these degrees due to different standards is crucial to the actual implementation work. It should be noted that this correspondence is usually not a one-to-one correspondence.

3.2 Coordination of Security Controls

The Chinese national standard GB/T 22239 puts forward protection requirements for different degrees of systems from three aspects: general management requirements, general technical requirements and expansion requirements, whereas IEC 63096 also proposes the security controls applicable to different life cycle stages for IACSs of NPPs with different security degrees. In fact, there is a strong overlap between the two. For example, the relevant requirements for the recruitment of personnel in the basic requirements for classified protection include "a) special departments or personnel shall be designated or authorized to be responsible for the recruitment of personnel; b) the identity, security background, professional qualifications shall be reviewed, and the technical skills possessed by the recruited personnel shall be assessed; c) confidentiality agreements should be signed with the recruited personnel and post responsibility agreements should be signed with the key positions", and the control measures on the "Prior to employment" of human resource security in IEC 63096 propose that "the organizational units in charge of I&C platform development, I&C system engineering and construction/ operation should have inventory of assets relevant to cybersecurity. They should develop a policy with associated procedures to review any potential individual when hired for a specific cybersecurity role, by consideration of the following:

Verification that the individual has the necessary cybersecurity awareness to perform the assigned role (e.g. based on performed trainings, existing certificates or respective job experience); Verification that the individual can be trusted to take on the role, especially if the role is critical for the organization". Therefore, , the coordination of security control measures between different standards should be fully considered before a power plant actually carries out cybersecurity work so as to better reduce time costs and investment costs.

3.3 Coordination Between Security and Nuclear Safety

Nuclear safety is of paramount importance to NPPs. Since IACSs of NPPs perform functions related to nuclear safety, especially those IACS important to safety, special attention should be paid to the coordination between nuclear safety and cybersecurity. The coordination should include at least two aspects. On the one hand, according to IEC 62859[12]], when the implementation of security controls may affect nuclear security functions, nuclear safety should be prioritized, and it is recommended that the above situations should be documented. On the other hand, it is suggested that the emergency response should ensure the nuclear safety when a cybersecurity incident occurs.

4 Conclusion

At present, we are carrying out the conversion of the IEC 63096 to the national standard, which is also the second one in the field of IACS cybersecurity of NPPs. With the deepening application of new technologies in NPPs such as cloud computing, big data,

Internet of Things, block chain and so on, IACSs of NPPs will face more known and unknown cybersecurity threats in the future. It's sure that international organizations, national standardization organizations and industry regulatory agencies will continue to develop relevant standards and regulatory requirements. The issues proposed in this paper that need to be considered in the process of coordinating the use of relevant standards by NPPs are also applicable to the coordination of future standards. Furthermore, if those issues are fully considered in the development process, the practicality of new standards will be further improved.

Acknowledgement

We thank all the reviewers.

Bibliography

[1] The China Nuclear Energy Association. China ranked second regarding the number of NPPs under construction and in operation, https://www.china-nea.cn/site/content/41339.html

[2] [1]KUSHNER, K., The Real Story of Stuxnet, https://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet

[3] Jianghai Li, Xiaojin Huang. Control System Security in Nuclear Power Plant[J]. Atomic Energy Science and Technology , 2012(46Suppl): 411-416.

[4] International Electrotechnical Commission. IEC 63096:2020 Nuclear power plants – instrumentation, control and electrical power systems – security controls[S]. Geneva: IEC, 2020.

[5] SAC/TC260, GB/T 22239-2019: Information Security Technology-Baseline for Classified Protection of Cybersecurity, Standardization Administration of The People's Republic Of China, 2019.

[6] SAC/TC260, GB/T 22240-2020: Information Security Technology-Classification Guide for Classified Protection of Cybersecurity, Standardization Administration of The People's Republic Of China, 2020.

[7] SAC/TC260, GB/T 25070-2019: Information Security Technology-Technical requirements of security design for classified protection of cybersecurity, Standardization Administration of The People's Republic Of China, 2019.

[8] SAC/TC260, GB/T 28448-2019: Information Security Technology-Evaluation requirements for classified protection of cybersecurity, Standardization Administration of The People's Republic Of China, 2019.

[9] SAC/TC260, GB/T 28449-2018: Information Security Technology-Testing and evaluation process guide for classified protection of cybersecurity, Standardization

Administration of The People's Republic Of China, 2018.

[10] SAC/TC40, GB/T 41241-2022: Management Requirements For Cybersecurity Of Industrial Control Systems In Nuclear Power Plant, Standardization Administration of The People's Republic Of China, 2022.

[11] IEC COMMITTEE 45, IEC 62645 Nuclear Power Plants – Instrumentation and Control Systems – Requirements for Security Programs for Computer-based Systems, IEC, 2014.

[12] IEC COMMITTEE 45, IEC 62859 Nuclear Power Plants – Instrumentation and Control Systems – Requirements for Coordinating Safety and Cybersecurity, IEC, 2016.

IT-Governance und Strategisches Informationsmanagement (ITG-SIM)

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1593

Agile Methoden für das IT-Controlling

Can Adam Albayrak[1], Andreas Gadatsch[2] und Benedikt Haag[3]

Abstract: Während sich die unternehmerische Arbeitswelt immer mehr in Richtung Agilität verschiebt, verharrt das IT-Controlling noch in alten, klassischen Strukturen. Diese Arbeit untersucht die Fragestellung, ob und inwieweit agile Ansätze im IT-Controlling eingesetzt werden können. Dieser Beitrag ist eine modifizierte Version des in der Zeitschrift „HMD Praxis der Wirtschaftsinformatik" (https://link.springer.com/article/10.1365/s40702-022-00837-0) erschienenen Artikels „Agiles IT-Controlling".

Keywords: IT-Controlling, IT-Projektmanagement, agil, Agilität, Scrum, Kanban

1 Einleitung

Seit vielen Jahren gehören agile Methoden zu beliebten Mitteln im IT-Projektmanagement. Besonders in Krisenzeiten, wie beispielsweise der Corona-Pandemie, offenbart sich die Stärke von Agilität in Unternehmen [SM20]. Auf Projektebene sind agile Strukturen inzwischen über die Softwareentwicklung hinaus weit verbreitet. Zu den wichtigsten agilen Methoden zählen Scrum, Kanban, Extreme Programming, Lean Software Development, Feature Driven Development, Adaptive System Development, Dynamic Systems Development Method sowie Crystal Clear [Tr21].

Im IT-Controlling werden in der Praxis überwiegend noch klassische Methoden des Projektmanagements eingesetzt [GKF17]. Agile Methoden finden bereits in vielen Managementkonzepten Anwendung und werden auch von der „Controlling-Disziplin" als relevantes Konzept wahrgenommen [JG20]. Die Digitalisierung führt dazu, dass zunehmend gefordert wird, dass sich die Controlling-Konzepte ändern müssen [Ke19]. Es stellt sich daher die Frage, ob die Methoden des IT-Controlling dieser Entwicklung Rechnung tragen müssen. Manche Autoren sehen darin sogar eine aktuelle Chance der Veränderung für das Controlling [SW18]. Dem „agilen Controlling" kann man sich dabei aus unterschiedlichen Perspektiven nähern. So ist offen, ob mit dem Begriff das Controlling agiler Organisationen/Systeme oder eine agile Arbeitsweise im Controlling selbst gemeint ist. In der vorliegenden Arbeit fokussieren die Autoren den Ansatz agiler

[1] Hochschule Harz, Friedrichstr. 57-59, 38855 Wernigerode, calbayrak@hs-harz.de
[2] Hochschule Bonn-Rhein-Sieg, Grantham-Allee 20, 53757 Sankt Augustin, andreas.gadatsch@h-brs.de
[3] Hochschule Bonn-Rhein-Sieg, Grantham-Allee 20, 53757 Sankt Augustin, benedikt.haag@h-brs.de

doi:10.18420/inf2022_138

Arbeitsweisen innerhalb des IT-Controllings.

Diese Arbeit untersucht die Frage, ob und wie im IT-Controlling bereits erfolgreich agile Methoden eingesetzt werden. Außerdem wird untersucht, inwieweit Elemente aus etablierten Frameworks, wie Scrum und Kanban sich ins IT-Controlling transferieren lassen.

Im Rahmen dieser Arbeit wird zunächst eine systematische Literaturanalyse durchgeführt, welche einen ausführlichen Einblick in den Stand der Forschung der Thematik gibt. Im Sommer 2021 wurde zudem eine Online-Umfrage unter Entscheidungsträgern und Fachexperten durchgeführt, welche Aussagen über den Einsatz von agilen IT-Controlling-Methoden in der Praxis ermöglichen soll. Die daraus gewonnen Erkenntnisse werden in Abschnitt 4 ausführlich thematisiert. Den Abschluss bilden eine Synthese sowie ein Zukunftsausblick.

2 Systematische Literaturanalyse

Die Autoren vertreten die These, dass die Literatur bisher noch nicht genauer den Einsatz agiler Ansätze im IT-Controlling behandelt haben. Um diese These zu überprüfen, wurde eine systematische Literaturanalyse unter Nutzung der Datenbanken SpringerLink, wiso, GVK, ACM Digital Library sowie ScienceDirect durchgeführt. Im Rahmen der Recherche wurden aus 1.473 Treffern insgesamt 22 themenrelevante Quellen identifiziert.

Die Analyse ergab, dass sich bisher nur ein geringer Teil der Literatur mit dem Thema *agiles IT-Controlling* auseinandersetzt. Gleichwohl wird die Bedeutung als sehr hoch eingeschätzt, da ein klassisches IT-Controlling in einer agilen Umgebung die angestrebte Agilität hemmt. Ein Großteil der Literatur setzt auf die Umsetzung von Agilität mittels geeigneter Frameworks, wie *Scrum* oder *Extreme Programming*, setzt. Während der Literaturtyp *Artikel* häufig spezielle agile Methoden und Frameworks behandelt, wird sich in *Fachbüchern* mit der Thematik lediglich oberflächlich auseinandergesetzt. Insgesamt zeigte sich, dass sich die Literatur häufig mit dem IT-Projektcontrolling auseinandersetzt und ein gesamtheitliches agiles IT-Controlling nicht vorkommt. Aufgrund von Platzmangel wird im Folgenden nur eine Auswahl empfehlenswerter Quellen für eine praktische Umsetzung von Agilität im IT-Controlling kurz präsentiert (s. Tab. 1).

Der Artikel „Kosten agiler Projekte steuern“ beschreibt die Vorteile und den Einsatz von klassischen *Burn Down Charts* für die Prognose anfallender Kosten in agilen Projekten (Gschmack 2021). In dem Buch „Projektcontrolling mit agilen Instrumenten“ werden die wichtigsten Instrumente des agilen Projektcontrollings charakterisiert und Umsetzungsempfehlungen anhand von Praxisbeispielen vorgestellt. Im Fokus steht dabei der Einsatz des Design-Thinking-Ansatzes [Kl21].

Kusay-Merkle zeigt in ihrem Buch „Agiles Projektmanagement im Berufsalltag“ die Einsatzmöglichkeiten agiler Methoden für die Planung und Umsetzung kleinerer und mittlerer Projekte. Damit richtet sie den Fokus auf Projekte, welche nicht die Größe eines „Vollzeitprojekts“ besitzen [Ku21].

Der Artikel „Digitale Innovationen agil bewerten“ setzt sich mit der Herausforderung der Risikobewertung agiler IT-Projekte auseinander. Das hier beschriebene *Fail Fast Framework* soll helfen, Entscheidungen über die Fortsetzung oder den Abbruch eines Projektes schneller und sicherer zu machen [MST20].

Preußig fokussiert in seinem Buch „Agiles Projektmanagement“ den Einsatz und die Umsetzung agiler Techniken in einem klassischen Projektumfeld. Hierzu zählen Projekte, welche die folgenden Merkmale aufweisen: Wasserfallprozess, Festpreisprojekt, Multiprojekt sowie Hierarchien [Pr20].

Autor(en)	Titel	Jahr	Verlag	Typ
Gschmack, S.	Kosten agiler Projekte steuern	2021	Controlling & Management Review	Artikel
Klein, A. (Hrsg.)	Projektcontrolling mit agilen Instrumenten	2021	Haufe-Lexware	Buch
Kusay-Merkle, U.	Agiles Projektmanagement im Berufsalltag	2021	Springer Gabler	Buch
Müller, A.; Schröder, H.; von Thienen, L.	Digitale Innovationen agil bewerten	2020	Controlling & Management Review	Artikel
Preußig, J.	Agiles Projektmanagement	2020	Haufe-Lexware	Buch

Tab. 1: Auswahl von aktuellen Quellen zum Thema „Agiles IT-Controlling“

3 Empirische Umfrage

Zur empirischen Absicherung wurden im Sommer 2021 Fach- und Führungskräfte in Unternehmen zum Thema *Agiles IT-Controlling* sowie zum Thema *Agiles Projektmanagement* befragt. Unter Berücksichtigung der vorherrschenden aktuellen äußeren Situation mit eingeschränkten Reisemöglichkeiten sowie der Tatsache, dass die Autoren bei den Befragten das sehr hohe und auch zum Teil andersartige Arbeitsaufkommen und die daraus resultierende zeitlich und inhaltlich intensivere Beanspruchung berücksichtigen wollten, erfolgte die Befragung ausschließlich online als Kurzumfrage. Die Durchführung einer ursprünglich geplanten ergänzenden Interviewreihe wurde aus den genannten Gründen aufgegeben.

3.1 Online-Umfrage

Den Befragten wurden online zum Thema *Agiles IT-Controlling* die Fragen aus Tabelle 2 sowie die Fragen aus Tabelle 3 zum Thema *Agiles Projektmanagement* gestellt:

	Fragenkatalog Agiles IT-Controlling
1	Setzen Sie im IT-Controlling agile Methoden ein?
2	Wenn JA: Seit wie vielen Jahren werden agile Methoden bei Ihnen im IT-Controlling eingesetzt?
3	Welche Vorteile haben oder hätten agile Methoden im IT-Controlling aus Ihrer Sicht? a) Bisherige Methoden waren nicht wirkungsvoll genug b) Schnellere Reaktion auf Veränderungen c) Lösungen sind stärker auf den (internen oder externen) Kunden/Anforderungssteller fokussiert d) Kontinuierliche Lieferung von Ergebnissen e) andere, nämlich …
4	Welche agilen Rahmenwerke (Frameworks) setzen Sie im Unternehmen oder in Ihrer Abteilung ein? a) Gar keine b) Scaled Agile Framework® (SAFe®) c) Kanban d) Scrum e) Design Thinking f) andere, nämlich …

Tab. 2: Fragenkatalog der Online-Befragung „Agiles IT-Controlling“

Für den Fall, dass bereits agile Methoden im IT-Controlling bzw. Projektmanagement eingesetzt werden, wurde darüber hinaus gefragt, wie die Erfahrungen mit Agilen Methoden beurteilt wurden. Die Antworten konnte auf einer Skala von (a) bis (e)

a) Ziele wurden nicht erreicht (keine wesentlichen Veränderungen)
b) Ziele wurden teilweise erreicht
c) Ziele wurden in großen Teilen erreicht
d) Ziele wurden überwiegend erreicht
e) Ziele wurden voll erreicht

angegeben werden. Abschließend wurde gefragt, welche Empfehlungen die Befragten jemandem geben würden, der agile Methoden erstmalig einsetzen möchte (Freitext von 200 Zeichen).

	Fragenkatalog Agiles Projektmanagement
1	Setzen Sie agile Methoden im Projektmanagement ein?
2	Wenn JA: Seit wie vielen Jahren werden agile Methoden bei Ihnen im Projektmanagement eingesetzt?
3	Vorteile haben oder hätten agile Methoden im Projektmanagement aus Ihrer Sicht?

	a)	Bisherige Methoden waren nicht wirkungsvoll genug
	b)	Schnelle Reaktion auf veränderte Anforderungen im Projekt
	c)	Lösungen sind stärker auf den (internen oder externen) Kunden/Anforderungssteller fokussiert
	d)	Kontinuierliche Lieferung von Projektergebnissen
	e)	andere, nämlich …
4	Welche agilen Rahmenwerke (Frameworks) setzen Sie im Projektmanagement ein?	
	a)	Gar keine
	b)	Scaled Agile Framework® (SAFe®)
	c)	Kanban
	d)	Scrum
	e)	Design Thinking
	f)	andere, nämlich …

Tab. 3: Fragenkatalog der Online-Befragung „Agiles Projektmanagement“

Befragt wurden die Empfänger eines elektronischen Infobriefes (Newsletter), Mitglieder der Fachgruppen *IT-Controlling* bzw. *Strategisches Informationsmanagement* der Gesellschaft für Informatik sowie einschlägige Gruppen in Sozialen Medien. Insgesamt gab es 34 qualifizierte Antworten, die sich wie in Tabelle 4 dargestellt verteilen.

Verteiler	Anzahl
elektronischer Infobrief	11 Teilnehmer
GI-Fachgruppen	18 Teilnehmer
Soziale Medien	5 Teilnehmer

Tab. 4: Verteilung der 34 qualifizierten Antworten

3.2 Ergebnisse der Befragung

Auch wenn die geringe Anzahl der Antworten aus erkenntnistheoretischer Sicht nur bedingt Schlussfolgerungen zulassen, konnten aber deutliche Indizien für einen bestimmten Trend beobachtet werden, den die Autoren am Ende dieses Abschnitts zusammenfassen. Zunächst werden auf die Ergebnisse der Befragung zum Agilen IT-Controlling eingegangen. In den insgesamt 34 qualifizierten Antworten gaben 13 Teilnehmer, also 38%, an, dass sie bereits Erfahrungen mit „Agilem IT-Controlling“ machen konnten. 62% gaben an, keinerlei Erfahrungen gemacht zu haben.

Die 13 Teilnehmer, die bereits Erfahrungen mit „Agilem IT-Controlling“ gemacht haben, antworteten auf die Frage seit wie vielen Jahren agile Methoden im IT-Controlling eingesetzt werden, ganz unterschiedlich: Während die meisten nur auf wenige Jahre Erfahrung

blicken können, gab es einen Teilnehmer, der bereits seit 9 Jahren Erfahrungen mit „Agilem IT-Controlling“ sammeln konnte. Die Details gehen aus Abbildung 1 hervor.

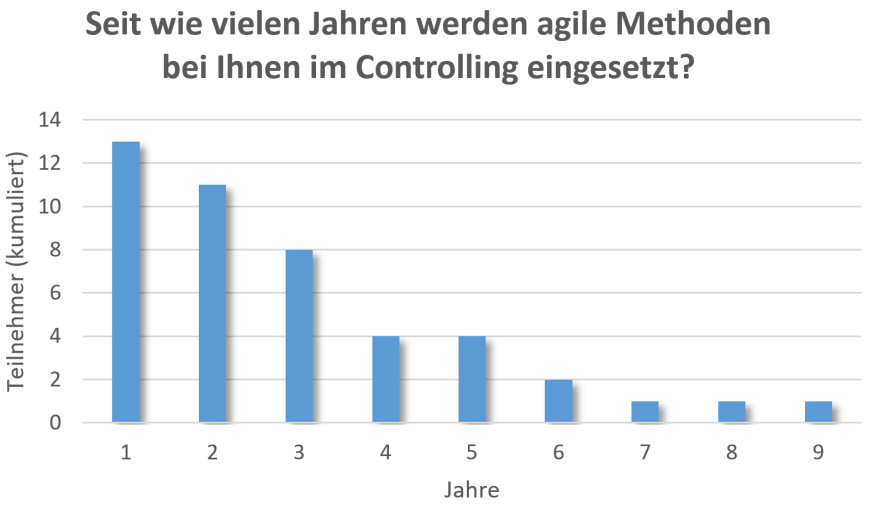


Abb. 1: Erfahrungen mit Agilem IT-Controlling (n=13)

Im weiteren Verlauf gaben 74% der 34 Teilnehmer an, dass agile Methoden im IT-Controlling Vorteile hätten. Der Hauptvorteil, der dabei in den 25 Antworten genannt wurde, ist die schnelle Reaktion auf Veränderungen sowie die stärkere Fokussierung der Lösung auf den Kunden/Anforderungssteller. Die Details gehen aus Abbildung 2 hervor.

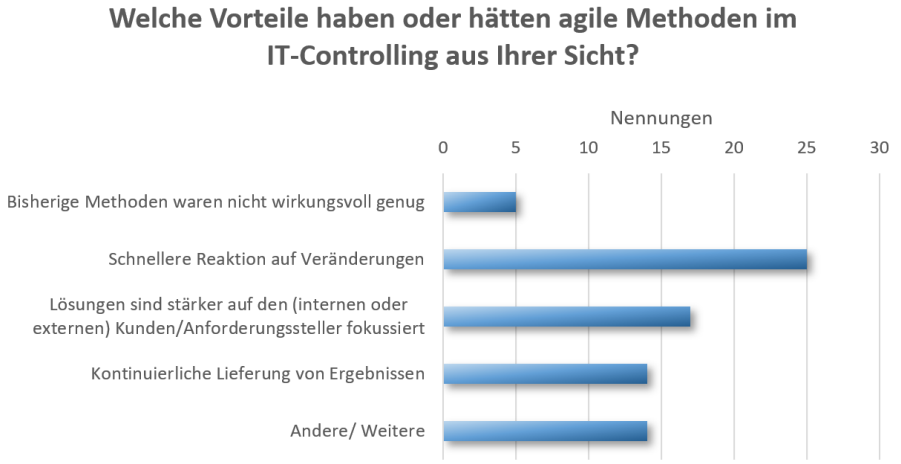


Abb. 2: Vorteile von Agilem IT-Controlling (n=25)

Die Umfrage ergab darüber hinaus, dass 82% agile Rahmenwerke einsetzen, vorzugsweise *Scrum* (s. Abb. 3).

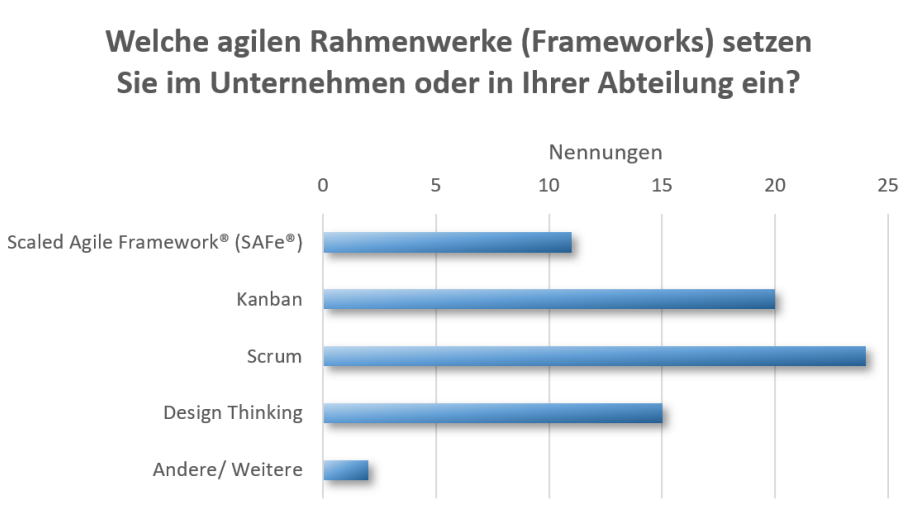


Abb. 3: Eingesetzte Rahmenwerke (n=28)

Im Folgenden werden die Ergebnisse der Befragung zum Agilen Projektmanagement dargestellt. Von den Teilnehmern gaben 85% an, dass sie bereits Erfahrungen mit Agilem Projektmanagement haben. Nur 15% der Teilnehmer haben bislang keinerlei Erfahrungen gemacht. Auf die Frage seit wie vielen Jahren agile Methoden im Projektmanagement

eingesetzt werden, gaben einige Teilnehmer sogar an, 10 Jahre und mehr solche Methoden einzusetzen. Die Details gehen aus Abbildung 4 hervor.

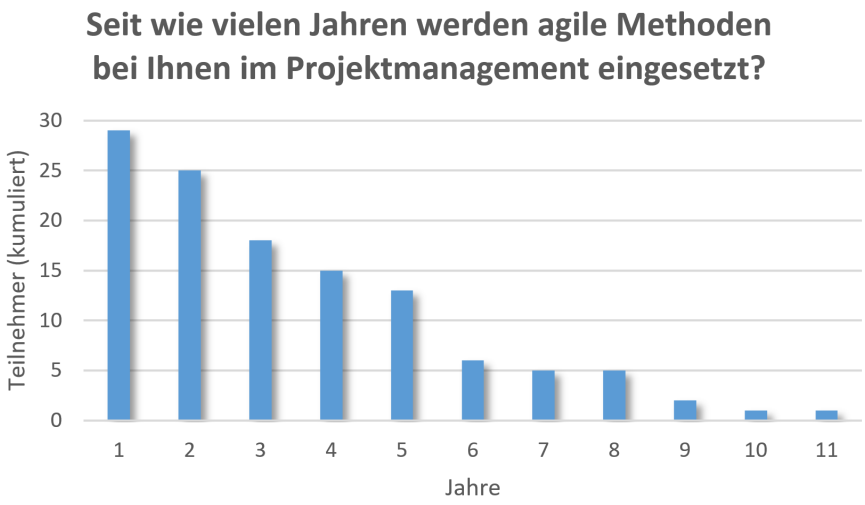


Abb. 4: Erfahrungen mit Agilem Projektmanagement (n=29)

Eine eindeutige Mehrheit vom 91% gab an, dass Agiles Projektmanagement Vorteile hätte. Der Hauptvorteil, der dabei von 31 Teilnehmern genannt wurde, ist ebenfalls die schnelle Reaktion auf Veränderungen sowie die stärkere Fokussierung der Lösung auf den Kunden/Anforderungssteller, dicht gefolgt von der kontinuierlichen Lieferung von Ergebnissen. Die Details gehen aus Abbildung 5 hervor.

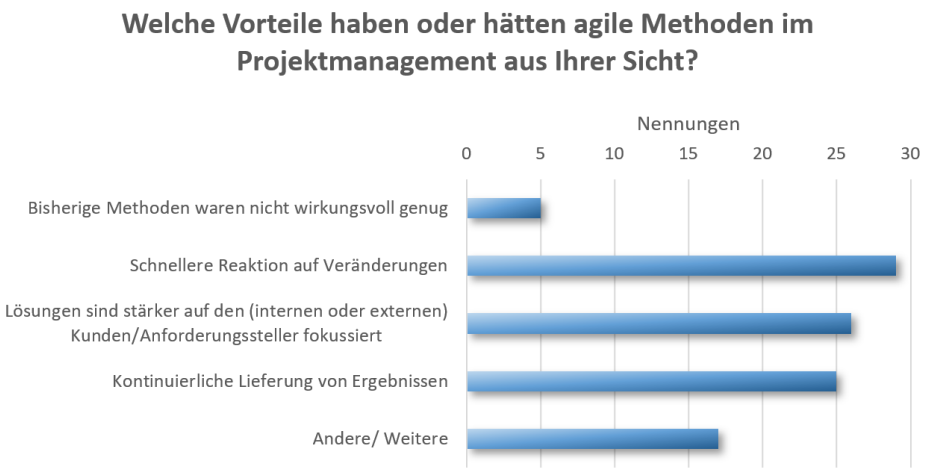


Abb. 5: Vorteile von Agilem Projektmanagement (n=31)

Die Umfrage ergab, dass 91% agile Rahmenwerke im Projektmanagement einsetzen, auch hier vorzugsweise *Scrum*, dicht gefolgt von *Kanban* (s. Abb. 6).

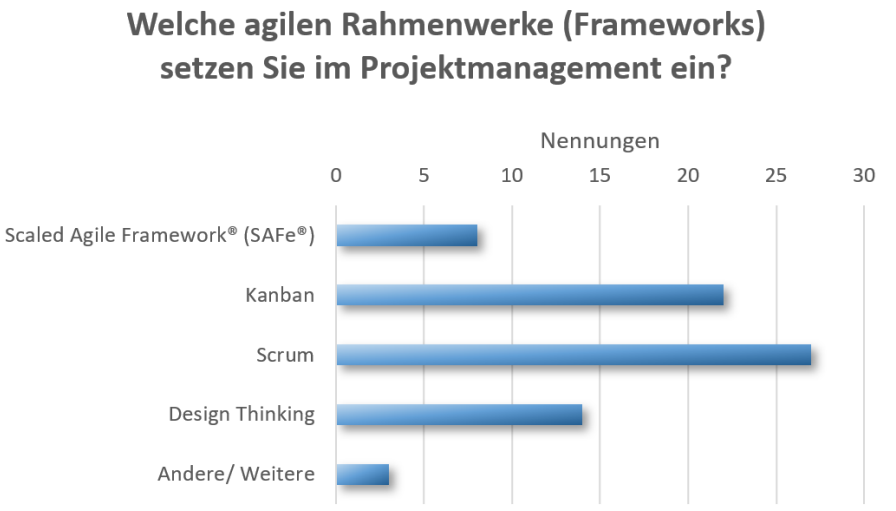


Abb. 6: Eingesetzte Rahmenwerke im Projektmanagement (n=31)

Zusammenfassend kann gesagt werden, dass die Teilnehmer der Befragung im Projektmanagement überwiegend (91%) agile Methoden eingesetzt haben und einsetzen. *Scrum* und *Kanban* sind hier die häufig eingesetzten Methoden.

Im IT-Controlling sieht das Bild anders aus: Agile Methoden sind hier weniger verbreitet (38%), wird von den meisten Teilnehmer aber als Werkzeug gesehen, um im IT-Controlling schnell auf Veränderungen reagieren zu können und darüber hinaus, um die Lösung des IT-Controllings stärker auf den Kunden bzw. Anforderungssteller zu fokussieren.

4 Allgemeine Eigenschaften eines Agilen IT-Controllings

Im Projektmanagement, insbesondere im Management von IT-Projekten, werden agile Methoden schon seit längerer Zeit eingesetzt. Bereits bevor Schwaber und Sutherland ihr agiles Vorgehensmodell zur Softwareentwicklung im Jahr 1995 auf der Konferenz *Object-Oriented Programming, Systems, Languages, and Applications* vorgestellt haben, haben Takeuchi und Nonaka im Jahr 1986 beschrieben, wie mit Methoden, die heute als agile Methoden bezeichnet werden, Produkte entwickelt werden können [TN86].

Überträgt man die in agilen Methoden für das Projektmanagement wünschenswerten Eigenschaften auf das IT-Controlling, so erhält man folgenden Anforderungskatalog für ein agiles IT-Controlling:

- Die höchste Priorität des IT-Controllings liegt auf der Zufriedenstellung des CIO und der IT-Manager durch zeitige und kontinuierliche Auslieferung von Ergebnistypen (Artefakten) des IT-Controllings.
- Statt den gesamten Entstehungszyklus der Artefakte vorherzusehen und zu planen erfolgt die Erstellung der Artefakte des IT-Controllings inkrementell, also in Teilschritten.
- Es werden nur so viele Artefakte des IT-Controllings erstellt, wie Zeit und Kosten es zulassen. Einfachheit ist bei der Erstellung wesentlich. Der Fluss der Arbeit wird optimiert.
- Die Priorisierung der zu erstellenden Artefakte des IT-Controllings erfolgt durch den CIO oder durch übrige IT-Manager.
- Die Arbeit des IT-Controllings wird für alle sichtbar gemacht.
- IT-Controller und Zulieferer arbeiten eng zusammen, häufig im persönlichen Gespräch von Angesicht zu Angesicht.
- Das IT-Controlling reflektiert in regelmäßigen Abständen, wie es effektiver werden kann und korrigiert seine Arbeitsweise gegebenenfalls.

Die in diesem Anforderungskatalog für ein agiles IT-Controlling beschriebenen *Artefakte* des IT-Controllings sind beispielsweise [Ga21]:

- Priorisiertes IT-Portfolio
- Bewertete Projektanträge (hinsichtlich Kosten, Risiken und Nutzen)
- Aufstellung eines abgestimmten IT-Budgets
- Bewertete IT-Projektlandschaft (hinsichtlich Kosten, Risiken und Nutzen)
- Bewertete IT-Systemlandschaft (hinsichtlich Kosten, Risiken und Nutzen)

- Ermittelter Wertbeitrag der IT am Unternehmenserfolg
- Beurteilte IT-Outsourcing-Maßnahmen (Chancen und Risiken)
- Bewertete Service-Level-Agreements mit IT-Dienstleistern und ggf. neugestaltete Service-Level-Agreements
- Vorbereite Make-or-Buy-Entscheidungen

5 Ansätze für ein Agiles IT-Controlling

In diesem Abschnitt generieren die Autoren Ansätze für ein Agiles IT-Controlling. Um die Herleitung nachvollziehen zu können, werden die aus Autorensicht wesentlichen Grundprinzipien der beiden häufig eingesetzten agilen Methoden Scrum und Kanban zuvor dargestellt.

5.1 Kanban

Gemäß *Kanban University* [Ka21] kann man mit der Kanban-Methode Arbeit managen. Es ist gemäß Kanban University eine Methode, um alle Arten von Wissensarbeit, zu managen. Die Kanban-Methode zu verwenden, bedeutet eine ganzheitliche Denkweise bei der Betrachtung der Services, mit einem Fokus auf deren Verbesserung aus Sicht ihrer Kunden, anzuwenden. Mit der Kanban-Methode wird Wissensarbeit und deren Weg durch die einzelnen Arbeitsschritte visualisiert. Kanban ist eine Managementmethode oder ein Managementansatz und kann auf bestehende Prozesse bzw. existierende Arbeitsweisen angewandt werden. Kanban ist dazu gedacht, die Arbeit besser zu managen und die Service-Erbringung soweit zu verbessern, dass die Erwartungen der Kunden zuverlässig erfüllt werden.

Die allgemeinen Praktiken von Kanban lauten verkürzt [Ka21]:

- **Visualisiere**: Eine gute Sichtbarmachung der Arbeit ist entscheidend für die effektive Zusammenarbeit und für das Erkennen von Verbesserungsmöglichkeiten. Wissensarbeit, die nur in den Köpfen stattfindet, ist unsichtbar. Die Visualisierung dieser Arbeit und des Flusses dieser Arbeit verbessert die Transparenz erheblich, wodurch zuvor unerkannte Probleme zum Vorschein treten.
- **Limitiere die parallele Arbeit** (das **WIP**): WIP (*Work in Progress*) stellt die Anzahl der parallel begonnen Arbeitsaufgaben dar. Durch Anwendung von Kanban wird erkannt, dass effektive Systeme mehr Fokus auf die Effizienz des Arbeitsflusses richten, als auf die Ressourcenauslastung. Wenn die Mitarbeiter einer Wertschöpfungskette zu 100% ausgelastet sind, gibt es naturgemäß keine Reserven in diesem System, wodurch der Arbeitsfluss bei der kleinsten Abweichung ins Stocken geraten kann. In der Wissensarbeit gibt es noch ein zusätzliches Problem: Kontextwechsel und Multi-Tasking. Je öfter man zwischen verschiedenen Themen umdenkt, desto ineffizienter wird man.

- **Steuere den Arbeitsfluss**: Das Ziel ist es, die Wertschöpfung so reibungslos, vorhersagbar und nachhaltig wie möglich zu gestalten. Wenn wir beginnen, den Fluss der Arbeit zu messen, erhalten wir überaus hilfreiche Informationen, um die Erwartungen der Auftraggeber zu managen, um Vorhersagen über Fertigstellungsdaten zu machen und Verbesserungen anzuregen.
- **Implementiere Feedbackschleifen**: Feedbackschleifen sind zwingend notwendig, um die Lieferfähigkeit des eigenen Service zu koordinieren und zu verbessern. Für den eigenen Kontext angepasste Rückkopplungsschleifen stärken die Lernbereitschaft der Organisation und deren Evolution durch Experimente. Übliche Feedback-Mechanismen in Kanban-Systemen sind Boards, Metriken und bestimmte wiederkehrende Meetings und Reviews.
- **Verbessere gemeinsam, entwickle experimentell weiter**: Kanban ist eine Methode, um kontinuierliche und inkrementelle Verbesserungen nachhaltig zu verfolgen.

5.2 Nutzung von Kanban-Board für agiles IT-Controlling

Betrachtet man den 7-Punkte-Anforderungskatalog aus Abschnitt 4, so stellt man fest, dass einige davon durch die Nutzung eines sog. *Kanban-Boards* unter Verwendung kollaborativer Tools [Mi22] unterstützt und gesteuert werden können. Insbesondere für verteilte Arbeiten wie *IT-Portfoliomanagement*, *IT-Budgetierung*, *IT-Projektcontrolling* u. a. lassen sich Kanban-Boards einsetzen. Ein Beispiel für die Nutzung eines Kanban-Boards im IT-Controlling unter Nutzung des Tools ist in (s. Abb. 7) dargestellt. Das beispielhafte Kanban-Board kann unter dem Link https://miro.com/app/board/o9J_lvHDTaU=/ betrachtet werden.

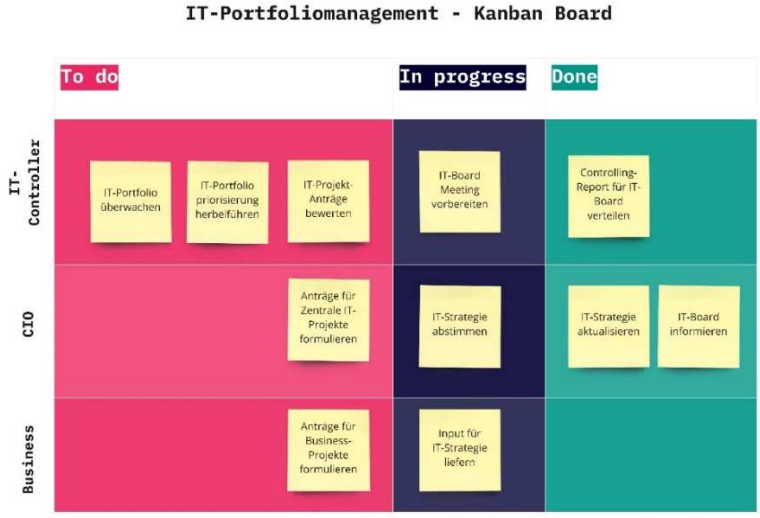


Abb. 7: Einfaches Beispiel für ein Kanban-Board im Agilen IT-Controlling (unter Verwendung des Tools Miro)

Darüber hinaus kann durch Anwendung der allgemeinen Kanban-Praktiken im IT-Controlling eine Annäherung an die Erfüllung des Anforderungskataloges erfolgen:

- Agile IT-Controller visualisieren ihre Arbeit (beispielsweise durch Kanban-Boards).
- Agile IT-Controller limitieren die parallele Arbeit durch konsequente Anwendung von WIP-Limits.
- Agile IT-Controller unterstützen nicht nur das IT-Management, sondern steuern auch ihren eigenen Arbeitsfluss durch geeignete Kennzahlen.
- Agile IT-Controller implementieren regelmäßige Feedbackschleifen über Kanban-Boards, über Metriken und über bestimmte wiederkehrende Meetings und Reviews mit dem Ziel, die Lieferfähigkeit der eigenen IT-Controlling-Dienste zu verbessern.
- Agile IT-Controller entwickeln die eigenen Methoden konsequent weiter.

5.3 Scrum

Die Scrum-Methode gehört zu den typischen und sehr häufig zum Einsatz kommenden Vertretern agiler Projektmanagementmethoden [SL18]. Die Methode geht bekanntlich auf Ken Schwaber und Jeff Sutherland zurück und wurde als Rahmenwerk zur Entwicklung und Erhaltung komplexer Produkte präsentiert [SS20]. Ihr Erfolg in der Praxis beruht darauf, dass klassische Methoden des Projektmanagements häufig scheitern, da sie die zu Beginn herrschende Unklarheit über das Projektziel, die notwendige Vorgehensweise und unbekannte zukünftige Anforderungen nicht berücksichtigen können.

Die zu Beginn des Projektes erstellten Pläne werden im Projektverlauf schnell von der Realität überholt. Ein wirkungsvolles Projektcontrolling ist daher nicht mehr möglich, da die Ausgangsbasis veraltet ist. Stattdessen wird bei Scrum darauf vertraut, dass gemischte Teams mit erfahrenen Mitgliedern Freiheiten erhalten, das Problem kreativ in Eigenverantwortung ohne starre Projektplanung zu lösen. Transparenz, Freiheit, zeitnahe Abstimmung des Teams und dezentrale Verantwortung ersetzen eine detaillierte zentrale Planung. Eine große Bedeutung haben hierbei die sogenannten *Sprints*, welche einzelne Softwarebausteine in vorgegebenen festen Zeitfenstern bearbeiten [HMN18]. Die Bewertung des Aufwands erfolgt mit *Story Points* auf der Basis von *User Storys*, welche die Anforderungen der geforderten Funktionalität beschreiben.

5.4 Bekannte und weniger bekannte Kennzahlen für agile Projekte

Vor dem Hintergrund der Kombination von agilen und klassischen Projekt-Methoden, aber auch bei rein agilen Projekten stellt sich die Frage, welche Kennzahlen aus dem klassischen Projektmanagement auch für agile Projekte sinnvoll und aussagekräftig sind sowie welche Kennzahlen überwiegend oder fast ausschließlich in der agilen Welt vorzufinden sind. Wir wollen hier nur eine Auswahl solcher Kennzahlen angeben.

Das klassische Informationsmanagement und das hierin eingebettete Projektcontrolling gehen von langfristig stabilen Anforderungen aus und streben nach Perfektion, während agile Konzepte eher Geschwindigkeit der Umsetzung mit kurzfristigen Änderungen im Fokus haben [Si16]. In der IT-Controlling-Literatur werden für das Projektcontrolling beispielsweise folgende Kennzahlen genannt:

- Anteil Aufwand der Kategorie X am Projektaufwand,
- Anzahl Probleme,
- Anteil übernommener Anforderungen,
- Fortschrittsgrad der Kategorie X,
- Fremdkräfteanteil,
- Ressourcenauslastung,
- Rentabilität des Projektes,
- Stabilität,
- Wirtschaftlichkeit des Projektes [Kü11].

Einige Kennzahlen können weitgehend unverändert übernommen werden (beispielsweise Fremdkräfteanteil, Ressourcenauslastung, Projektrentabilität), andere dagegen (z. B. Fortschrittsgrad) müssen angepasst werden.

Die fehlende stabile Planung stellt das klassische Projektcontrolling vor eine Herausforderung, da der klassische Plan-Ist-Vergleich oder Soll-Ist-Vergleich von Projektkennzahlen nicht mehr direkt übernommen werden kann. Ein möglicher Ansatz besteht darin, Projektkennzahlen mit Fokus auf die Performance von Sprints zu entwickeln. Weitere

Möglichkeiten bestehen darin, die *Story Points* in Euro zu bewerten und in Form von Kennzahlen zu analysieren.

Kennzahlen, die überwiegend oder nahezu ausschließlich in agilen Projekten Verwendung finden sind:

- *Aufwand* bzw. *Zeit für MVP*: Aus der Lean-Start-Up-Szene stammt das Konzept des *Minimum Viable Product* (MVP) bei dem ein *MVP* diejenige Produktversion ist, welche mit minimalem Aufwand einen Mehrwert für den Kunden liefert. Schätzungen, wie viele Personentage oder *Story Points* für ein MVP benötigt werden bzw. wie lange es dauert, bis das MVP beim Kunden Nutzen stiften kann, sind Kennzahlen, die überwiegend oder nahezu ausschließlich in agilen Projekten Verwendung finden.

- Finanzielle Kennzahlen wie
  - der *Geschäftswert*, also die Beantwortung der Frage „Welchen Wert hat eine bestimmte Produkteigenschaft für den Kunden?“
  - *oder die Verzögerungskosten*, also die Beantwortung der Frage „Wie hoch sind die Kosten bei Verzögerung des Projektes?“,

  sind Kennzahlen, die sowohl bei klassischen als auch bei agilen Projekten Verwendung finden.

- Leistungskennzahlen, die ausschließlich bei agilen Projekten Verwendung finden, sind:
  - *WIP* (*work in progress*), also die Anzahl der Aufgaben, an denen gerade gearbeitet wird inklusive der für Kanban typischen WIP-Begrenzungen (*WIP limits*),
  - die *Durchlaufzeit* (*lead time*), die in Kanban angibt, wie lange es dauert, eine Funktion zu entwickeln,
  - der *Velocity-Faktor* als Maßeinheit für die Geschwindigkeit eines Scrum-Teams, welche misst, wie viele *Story Points* ein Scrum-Team pro Sprint durchschnittlich erledigen kann,
  - sowie visuelle Darstellungen des Projektfortschritts und darüber, wie viele Anforderungen sich zu welchem Zeitpunkt in welchem Umsetzungszustand befinden, in Kanban durch *kumulative Flussdiagramme* (*Cumulative Flow Diagram*, CFD) oder zum Teil in Scrum auch durch sogenannte *Burndown Charts*.
  - Mitarbeiterkennzahlen wie *Mitarbeiterzufriedenheit*, *Mitarbeiterfluktuationsrate* im Projekt, *Krankenstand* bzw. *Gesundheitsstand* im Projekt sind Kennzahlen, die sowohl bei klassischen als auch bei agilen Projekten Verwendung finden.

- Kundenkennzahlen wie
  - *Kundenzufriedenheit*, ermittelt durch eine neutrale Instanz, die eine Kundenumfrage durchführt
  - sowie vom Kunden bereits *abgenommene Inkremente* im Projekt

sind Kennzahlen, die sowohl bei klassischen bzw. bei agilen Projekten Verwendung finden.

- Kennzahlen, die den Fluss der Arbeit beschreiben wie
  - *Meetingtreue* (Quotient aus den tatsächlich durchgeführten Besprechungen und den geplanten Besprechungen)
  - *Meetingzeitüberschreitungen* in Minuten
  - *Wartezeiten* in Tagen bei blockierten Aufgaben

  sind typische Kennzahlen in agilen Projekten.

5.5 Agile Erstellung der IT-Budgetplanung

Traditionelle Budgetplanung, so auch die Planung der IT-Budgets, fußt meist auf produktbezogene Absatz- und Umsatzziele sowie oftmals auf Fortschreibung der bisherigen Budgetpläne. Der Fokus ist zudem meist auf den Zeitraum eines Geschäftsjahres gerichtet. Der Aufwand für die Erstellung der Budgetpläne ist häufig hoch. Obwohl an der verwendeten Budgetierungspraxis sowohl in der Literatur als auch in der Praxis vielfach Kritik geäußert wurde, hat sich nach Einschätzung der Autoren hieran in den Unternehmen nur wenig verändert. Ansätze wie *Beyond Budgeting*, *Advanced Budgeting*, *Better Budgeting* (s. Abb. 8) sollen mehr Dynamik zulassen und die Missstände des Planungsprozesses beseitigen.

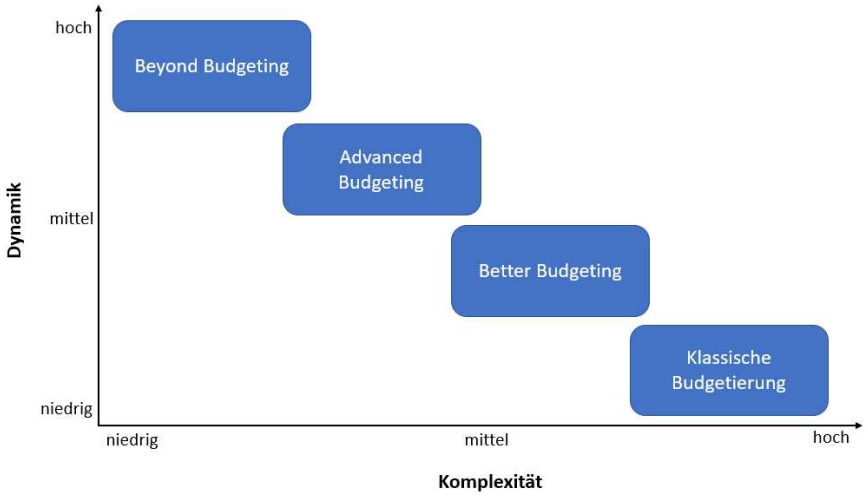


Abb. 8: Verschiedene Budgetierungsansätze [HGS19]

Die Basis beim *Beyond Budgeting* bilden 12 als vorbildlich betrachtete Management- und Controlling-Prinzipien, die sich konsequent an den Erfordernissen des Marktes ausrichten,

aber untereinander abhängig sind. Relative Ziele treten an die Stelle von fixierten Budgetgrößen. Mit seinen Prinzipien wie

- Teams die Freiheit und den Raum zum Handeln geben, ohne Mikro-Management von oben zu betreiben,
- ein schlankes Netzwerk aus ergebnisverantwortlichen Teams erschaffen, keine zentralistische, funktional geteilte Pyramide,
- relative Ziele für kontinuierliche Verbesserung setzen, keine fixierten Leistungsverträge verhandeln,

passt der Ansatz des *Beyond Budgeting* am ehestens zu einer „Agilen Budgetierung".

5.6 Agiler Verteilungsprozess für IT-Budgets

Neben dem Planungsprozess der IT-Budgets muss auch der Prozess von der Projektidee zum IT-Projekt betrachtet werden, bei dem Vorhaben mit IT-Budgets zur Umsetzung ausgestattet werden müssen. Im klassischen Verteilungsprozess werden Projektanträge bewertet und nach Genehmigung mit dem kompletten Budget ausgestattet. Helmke und Uebel empfehlen hierzu die Etablierung eines Gremiums zur Freigabe von IT-Budgets nach erfolgter Bewertung und Priorisierung der Projektanträge [HU16].

Beim Einsatz von agilen Projektmanagementmethoden passt dieses Vorgehen der Genehmigung von IT-Budgets nicht mehr, da bei agilen Projekten insbesondere Kosten und Zeit vorgegeben sind und der Umfang der IT-Lösung variiert (s. Abb. 9). Darüber hinaus besteht die Möglichkeit, Budgets pro „Bauabschnitt", also pro Inkrement oder einer Sammlung von Inkrementen, nach erfolgter (kontinuierlicher Auslieferung) freizugegeben.

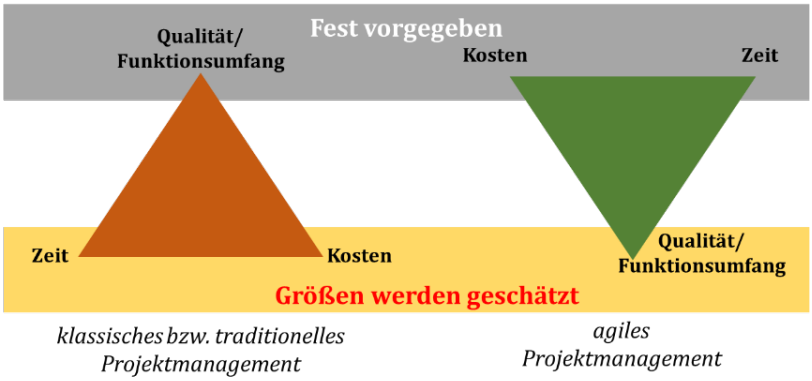


Abb. 9: Fest vorgegebene Größen im IT-Projektmanagement

Daher wird von den Autoren empfohlen, lediglich die Planung der IT-Budgets für die nächsten beiden(!) „Bauabschnitte" im Detail vorzunehmen, wodurch die Berücksichtigung aktueller Ereignisse in den jeweiligen Anforderungen möglich wäre und Mehr- oder Minderaufwände relativ zeitnah identifiziert werden können. Erfolgt zudem eine

Aktualisierung der prognostizierten Rentabilität, kann eine Repriorisierung auf Sprint-Ebene erfolgen. Entsprechende Ansätze wurde von einem der Autoren bereits prototypisch im Unternehmen erprobt und evaluiert [CA18].

6 Der Faktor Mensch im Agilen IT-Controlling

Neben dem Einsatz agiler Methoden gehört auch eine agile Denk- und Handlungsweise zu den Voraussetzungen für einen erfolgreichen Wandel hin zur Agilität. Die Relevanz zeigt sich insbesondere bei der Betrachtung der häufigsten Gründe für ein Scheitern. Hierzu gehören Probleme mit der Organisationskultur, Widerstand gegen Veränderung sowie mangelnde Unterstützung und Fähigkeiten [Di21]. Gleichzeitig beschreibt Denning in seinem Buch „The Age of Agile“, dass die Veränderung der Organisationskultur zu den größten Herausforderungen bei der Implementierung eines agilen Managements zählt [De18]. Es stellt sich daher die Frage, wie sich die Rolle des IT-Controllers in Zukunft verändern muss und wie die „richtige Denkweise“ aussehen muss. Die Beantwortung dieser Fragestellung sollte in zukünftigen Forschungsarbeiten weiter untersucht werden.

Eine weitere Besonderheit ergibt sich aus der Rolle des IT-Controllers in einem Unternehmen an sich: Während in Großunternehmen durchaus ein ganzes Team mit den Aufgaben des IT-Controlling betraut ist, kann in vielen Unternehmen beobachtet werden, dass das IT-Controlling nur aus einer einzigen Stelle besteht. Mancherorts stellt die Tätigkeiten des IT-Controllings sogar nur einen Teil der Aufgaben dar, die ein Mitarbeiter auf seiner Stelle ausführt. Dies wiederum passt überhaupt nicht zu den agilen Ansätzen, die durch Arbeit in selbstorganisierenden Teams geprägt ist und in denen die Expertise auf die Teammitglieder verteilt ist. Auch die Beantwortung dieser Fragestellung sollte in zukünftigen Forschungsarbeiten weiter untersucht werden.

7 Fazit

Es zeigt sich, dass das agile IT-Umfeld zunehmend auch Auswirkung auf das IT-Controlling hat. Während in der Praxis häufig noch auf klassische Strukturen eingesetzt werden, entstehen in der Literatur bereits erste Ansätze für ein agiles IT-Controlling, jedoch bislang ohne ganzheitliches Konzept. Zudem konnte herausgefunden werden, dass die Rolle und die Denkweise eines IT-Controllers erheblichen Einfluss auf den Erfolg der agilen Transformation hat. Auf Basis der in diesem Artikel gewonnen Erkenntnisse, können in folgenden Arbeiten detaillierter auf einzelne Aspekte eingegangen werden. So können bereits vorhandene Konzepte, wie die im SAFe-Framework enthaltenen Bereiche Lean Budgeting und Lean Portfoliomanagement sowie Aspekte des Flight-Level Konzepts, aufgegriffen und auf ein agiles IT-Controlling übertragen werden.

Literaturverzeichnis

[CA18] Certa, Sandra Stefanie; Albayrak, Can Adam: Eine hybride Vorgehensweise zur IT-Projektportfolioplanung. In (Mikusz, M.; Volland, A.; Engstler, M.; Fazal-Baqaie, M.; Hanser, E.; Linssen, O. Hrsg.): Projektmanagement und Vorgehensmodelle 2018. Der Einfluss der Digitalisierung auf Projektmanagementmethoden und Entwicklungsprozesse, Gesellschaft für Informatik, Bonn, 2018.

[De18] Denning, S.: The Age of Agile. How Smart Companies Are Transforming the Way Work Gets Done. AMACOM, 2018.

[Di21] Digital.ai (Hrsg.): The 15th State of Agile Report. Agile adoption accelerates acros the enterprise, 2021.

[Ga21] Gadatsch, A.: IT-Controlling. 2. Aufl., Springer, Wiesbaden, 2021.

[GKM16] Gadatsch, A.; Komus, A.; Mendling, J.: BPM-Compass 2016; Eine wissenschaftliche Studie der Hochschule Koblenz, Hochschule Bonn-Rhein-Sieg und der Wirtschaftsuniversität Wien in Zusammenarbeit mit der Gesellschaft für Prozessmanagement e.V. www.project-and-process.net/BPM-Compass, 2016.

[GKF17] Gadatsch, A.; Kütz, M.; Freitag, S.: Ergebnisse der 5. Umfrage zum Stand des IT-Controllings im deutschsprachigen Raum. Schriftenreihe des Fachbereiches Wirtschaft Sankt Augustin, Hochschule Bonn-Rhein-Sieg, 2017.

[Gs21] Gschmack, S.: Kosten agiler Projekte steuern. Controlling & Management Review 65:56–59, 2021.

[HMN18] Hansen, H-R.; Mendling, J.; Neumann, G.: Wirtschaftsinformatik. 12. Aufl., De Gruyter Oldenbourg, Berlin, 2018.

[HU16] Helmke, S.; Uebel, M. (Hrsg.): Managementorientiertes IT-Controlling und IT-Governance. 2. Aufl., Springer Fachmedien Wiesbaden, Wiesbaden, 2016.

[HGS19] Horváth, P.; Gleich, R.; Seiter, M.: Controlling. 14. Aufl., Franz Vahlen, 2019.

[JG20] Janka, M.; Günther, T.: Controlling für agiles Management. Controlling & Management Review 64:24–33, 2020.

[Ka21] Kanban University (Hrsg.): Der offizielle Leitfaden zur Kanban-Methode, 2021.

[Ke19] Kenfenheuer, K.: Neue Controlling-Konzepte gefragt. Controlling & Management Review 63:32–37, 2019.

[Kl21] Klein, A. (Hrsg.): Projektcontrolling mit agilen Instrumenten. Grundlagen, Werkzeuge, Praxisbeispiele, 1. Aufl., Haufe-Lexware, Freiburg, 2021.

[Ku21] Kusay-Merkle, U.: Agiles Projektmanagement im Berufsalltag. Für mittlere und kleine Projekte, 2. Aufl. Springer, Berlin, Heidelberg, 2021.

[Kü11] Kütz, M.: Kennzahlen in der IT; Werkzeuge für Controlling und Management. 4. Aufl. dpunkt.verlag, Heidelberg, 2011.

[Mi22] Miro (2022): Unternehmenswebseite. Online verfügbar unter www.miro.com, zuletzt geprüft am 13.05.2022.

[MST20] Müller, A.; Schröder, H.; Thienen, L. von: Digitale Innovationen agil bewerten. Controlling & Management Review 64:50–55, 2020.

[Pr20] Preußig, J.: Agiles Projektmanagement. Agilität und Scrum im klassischen Projektumfeld, 2. Aufl., Haufe-Lexware, Freiburg, 2020.

[SW18] Scharner-Wolff, P., Witte, E.: Agiles Controlling. Veränderung als Chance, Controlling & Management Review 62:24–33, 2018.

[SL18] Schelle, H.; Linssen, O.: Projekte zum Erfolg führen. Projektmanagement systematisch und kompakt, 8. Aufl., Beck im dtv, München, 2018.

[SS20] Schwaber, K.; Sutherland, J.: The Scrum Guide. Online verfügbar unter https://www.scrum.org/resources/scrum-guide, zuletzt geprüft am 13.05.2022, 2020.

[Si16] Sieber, R.: Die IT der Zukunft hat zwei Geschwindigkeiten: agil und stabil. https://different-thinking.de/die-der-zukunft-hat-zwei-geschwindigkeiten-agil-und-stabil/. Zugegriffen: 15. November 2021, 2016.

[SM20] Spreckelsen, J. von; Mitter, A.: Wie agil ist Ihre Organisation? Agile Pulse 2020 – die Agilitätsstudie von BearingPoint, v. BearingPoint. Frankfurt am Main, 2020.

[TN86] Takeuchi H.; Nonaka, I.: The New New Product Development Game. Harvard Business Review:137–146, 1986.

[Tr21] Trivedi, D.: Agile Methodologies. In: International Journal of Computer Science & Communication 12, S. 91–100, 2021.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1613

Beschreibung von IT-Services für Kunden: Vergleich und Konsolidierung der Attribute auf Basis einer Metastudie

Sebastian Floerecke und Franz Lehner[1]

Abstract: IT-Service-Beschreibungen müssen einheitlich und kundenorientiert gestaltet sein und dabei alle Informationen umfassen, die Kunden für Entscheidungen bezüglich sämtlicher Service-Request-Arten wie Neubestellung, Änderung und Kündigung benötigen. Für die Festlegung und Befüllung der notwendigen Attribute liegen in der wissenschaftlichen Literatur zwar mehrere Ansätze vor, ein kritischer Vergleich, eine Zusammenführung und damit die Konsolidierung der bestehenden Konzepte ist aber noch ausstehend. Dieser Umstand weist einerseits auf einen Forschungsbedarf hin, andererseits werden Unternehmen bei der Beschreibung von IT-Services gegenwärtig nur unzureichend unterstützt. Vor diesem Hintergrund wurde eine vergleichende und integrative Metastudie durchgeführt, bei der mittels einer strukturierten Literaturrecherche die bestehenden Ansätze zunächst ermittelt und anschließend analysiert und konsolidiert wurden. Zentrales Ergebnis ist ein auf Basis von 24 Studien konsolidierter Katalog mit 69 Attributen. Die vorhandenen Beschreibungsansätze weisen, sowohl die Anzahl als auch die Ausprägung der Attribute betreffend, deutliche Unterschiede auf. Ein gemeinsamer Kern an Attributen lässt sich aktuell nicht erkennen. Der geschaffene Attributkatalog liefert die Grundlage für weitere Untersuchungen wie die Überprüfung der Vollständigkeit. Praktiker erhalten eine Referenz für die Erstellung bzw. Überarbeitung von IT-Service-Katalogen.

Keywords: IT-Service-Management (ITSM), IT-Service-Katalog, Business-Service, IT-Service, Service-Beschreibung, Attribute, Literaturrecherche, Metastudie

1 Problemstellung und Zielsetzung

Der IT-Service-Katalog eines Unternehmens ist eine strukturierte Aufstellung und Beschreibung, die transparent machen soll, welche IT-Services zu welchen Merkmalsausprägungen und zu welchem Preis von Personen der eigenen und/oder einer anderen Organisation bezogen werden können [Ax19; MA18]. IT-Service-Kataloge schaffen damit eine gemeinsame Kommunikationsgrundlage für Anbieter und Nachfrager und bilden die Basis für zahlreiche IT-Service-Management (ITSM)-Prozesse wie dem Configuration-, Change-, Incident- und Service-Request-Management sowie für die innerbetriebliche Leistungsverrechnung und (externe) Auftragsabrechnung [MA22; SH18]. Zur Unterstützung und Umsetzung der verschiedenen ITSM-Prozesse inklusive dem IT-Service-Katalog-Management setzen Großunternehmen und reifere KMUs in der Regel ITSM-Software-Suiten wie EasyVista, ServiceNow und USU ein [Fl22].

[1] Universität Passau, Lehrstuhl für Wirtschaftsinformatik mit Schwerpunkt Informations- und IT-Service-Management, Innstraße 43, 94032 Passau, {sebastian.floerecke, franz.lehner}@uni-passau.de

doi:10.18420/inf2022_139

Eine 2021 mit über 500 IT-Entscheidern in der DACH-Region vom Marktforschungsunternehmen IDG Research Services durchgeführte Internetbefragung [ID21] ergab, dass etwa die Hälfte der Unternehmen mit ihrem ITSM nicht zufrieden ist. Diese Unzufriedenheit betrifft, wie weitere Studien offenlegen, im besonderen Maße das IT-Service-Katalog-Management. Die gemeinsame Online-Umfrage des Softwareherstellers USU und des Trainingsanbieters SERVIEW [BL19] im Jahr 2019 zeigte beispielsweise, dass nur 14 % der 140 mitwirkenden Unternehmen über einen IT-Service-Katalog mit einer genauen Auflistung und Beschreibung aller angebotenen IT-Services verfügten. 56 % besaßen einen unvollständigen und 30 % überhaupt keinen IT-Service-Katalog. Nach einer bereits 2016 von der ITSM-Consulting-Firma Evergreen durchgeführten Untersuchung [Ca16] betrachteten lediglich 57 % der rund 100 beteiligten Unternehmen, die versuchten, einen IT-Service-Katalog zu gestalten, das Vorhaben als Erfolg. 12 % sahen das Projekt gar als gänzlich gescheitert an. Als größte Schwierigkeit werden vielfach die Definition und die Beschreibung der einzelnen IT-Services genannt.

Vielen Unternehmen bereitet es offenkundig erhebliche Schwierigkeiten, einen vollständigen, adäquat strukturierten und beschriebenen IT-Service-Katalog zu erstellen [Fl21b; MA22; NDB16; Sc17a]. Die Gründe für das häufige Scheitern bzw. die mangelhaften Resultate sind vielfältig. Zum einen mangelt es an bewährten, öffentlich zugänglichen Vorlagen und die wenigen vorhandenen sind unvollständig, unzureichend beschrieben und damit weitgehend unpraktikabel [Fl21c; MA22; NJ15; RBK08; Sc17a]. Zum anderen existieren in der Fachliteratur lediglich vereinzelte, grobgranulare Vorschläge für ein Vorgehensmodell (z.B. [MB09; SS16]). Die Forschung hat sich bislang auf die Identifizierung, Erläuterung und Diskussion der erforderlichen Komponenten eines IT-Service-Katalogs konzentriert, ohne den Entwicklungsweg genauer zu spezifizieren [SH18].

Eine Schlüsselaktivität bei der Anfertigung eines IT-Service-Katalogs ist die Festlegung seiner Grundstruktur durch die Definition mehrstufiger, miteinander verbundener Strukturierungsebenen [MA22; MA18; NDB16; Sc15]. Eine für Kunden verständliche Grundstruktur ist sowohl in Dokumentenform als auch für Self-Service-Portale unabdingbar [Fl21c; Sc17b]. Des Weiteren ist es erforderlich, die einzelnen einsortierten IT-Services anhand von geeigneten Attributen einheitlich und kundenzentriert zu beschreiben [Fl21c; MA22; MA19; NJ15; RBK08]. Die Beschreibungen müssen dabei sämtliche Informationen enthalten, die Kunden für Entscheidungen bezüglich aller Service-Request-Arten – Neubestellung, Änderungen jeglicher Art und Kündigung – benötigen. Für die Festlegung und Befüllung der notwendigen Attribute existieren in der Fachliteratur zwar erste Ansätze [Fl21a; PZ19; RBK08], ein kritischer Vergleich, eine Zusammenführung und damit die Konsolidierung der bestehenden Konzepte fehlen jedoch [MA22; MA18].

Zur Verbesserung dieser gleichermaßen für die Praxis und die Wissenschaft unbefriedigenden Situation besteht das Ziel dieses Beitrags in einer Konsolidierung der bisher vorgeschlagenen Beschreibungsattribute mithilfe einer Metastudie. Mit einer strukturierten Literaturrecherche werden die in der wissenschaftlichen Literatur vorliegenden Ansätze zunächst ermittelt, anschließend mit einem konzeptzentrierten Vorgehen analysiert und die Attribute in einer Konzeptmatrix aggregierten Konzepten zugeordnet. Neben den

Attributen umfassen die betrachteten Konzepte die Forschungsmethode, das Veröffentlichungsjahr, die Veröffentlichungssprache und die Publikationsart. Der Beitrag ist folgendermaßen aufgebaut: Nachdem im Kapitel 1 die Problemstellung und die Zielsetzung dieser Untersuchung erläutert wurden, befasst sich Kapitel 2 mit den Grundlagen von IT-Services, ITSM und IT-Service-Katalogen. Kapitel 3 beschreibt das Forschungsdesign. Im Kapitel 4 werden die Ergebnisse vorgestellt und im Kapitel 5 kritisch reflektiert.

2 Grundlagen von IT-Services, ITSM und IT-Service-Katalogen

Der Begriff IT-Service wird in der wissenschaftlichen Literatur und der Praxis nicht einheitlich verstanden und verwendet. Die darunter subsumierten Inhalte sind zudem einem kontinuierlichen Wandel unterworfen [RR18; Sc17a]. Dem breiten Verständnis des IT Service Management Forum (itSMF) folgend, besteht ein IT-Service aus einer Kombination von Personen, Prozessen und IT, ist über Service-Level-Agreements (SLAs) definiert und unterstützt die Geschäftsprozesse eines oder mehrerer Kunden. IT-Services werden ähnlich einem Produkt als abgeschlossene Einheit angeboten [IT07]. Die einzelnen Leistungsbestandteile lassen sich nach den erforderlichen Fähigkeiten zur Umsetzung bzw. dem Digitalisierungsgrad unterteilen [Ga22]. So enthalten IT-Services für gewöhnlich manuelle Komponenten wie Beratung, Customizing, Aufbau, Support und Abbau sowie digitale und damit automatisierbare Komponenten wie Betrieb, Monitoring und Daten-Backup einer Applikation – der IT-Service im engeren Sinn [WTT14].

Zur Umsetzung des ITSM bieten die Literatur und die Praxis Rahmenwerke und Standards. Hierzu zählen unter anderem COBIT, ISO/IEC 20000 sowie das Microsoft-Operations-Framework (MOF). Am weitesten verbreitet und seit Längerem als De-facto-Standard gilt die Information Technology Infrastructure Library (ITIL) [MA19; Ti22]. ITIL definiert ITSM als „*[e]ine Reihe spezialisierter Fähigkeiten einer Organisation zur Generierung eines Werts für Kunden in Form von Services*“ [Ax19]. Das übergeordnete Ziel von ITSM besteht demnach darin, die vorhandenen Fähigkeiten und Ressourcen so zu managen, dass die Gestaltung und die Erbringung der IT-Services optimal auf die Anforderungen der Kunden ausgerichtet ist [BZ14]. Da ITIL Unternehmen größen- und branchenübergreifend abdecken möchte, werden nicht einzelne Arbeitsschritte, sondern Erfahrungen und Bewährtes auf relativ hoher Abstraktionsebene beschrieben [KB17].

Die aktuelle ITIL-Version (v4) umfasst 34 ITSM-Prozesse. Einer davon ist das Service-Katalog-Management, dessen Hauptziel in der Bereitstellung und der Pflege eines vollständigen, adäquat strukturierten und beschriebenen IT-Service-Katalogs besteht [Ax19]. Der IT-Service-Katalog eines Unternehmens ist eine Aufstellung und Beschreibung, die transparent machen soll, welche IT-Services zu welchen Merkmalsausprägungen und zu welchem Preis von Personen der eigenen (interne Kunden) und/oder einer anderen Organisation (externe Kunden) bezogen werden können [Ax19; MA18]. IT-Service-Kataloge schaffen eine gemeinsame Kommunikationsgrundlage für Anbieter und Nachfrager und fungieren als Basis zahlreicher ITSM-Prozesse wie dem Service-Request-, Incident-,

Configuration- und Change-Management. Sie dienen zudem der innerbetrieblichen Leistungsverrechnung und externen Auftragsabrechnung [MA22; SH18]. Im IT-Service-Katalog können sowohl eher atomare IT-Services wie einzelne Applikationen oder Geräte (Bildschirm, Drucker, Smartphone, etc.) als auch Service-Pakete, bestehend aus integrierten und gegebenenfalls optionalen Komponenten, enthalten sein. Ein Beispiel eines Service-Pakets wäre ein kompletter Büroarbeitsplatz, der sich aus Desktop-PC mit Betriebssystem, Bildschirm, Standard-Office-Software, Spezial-Applikationen und Telefon zusammensetzt. Dieser Bündelansatz erleichtert Kunden die Bestellung technisch kompatibler und fachlich sinnvoller IT-Service-Kombinationen. Werden IT-Services auch externen Kunden angeboten, lässt es sich dadurch gegenüber Wettbewerbern, deren Portfolios keine Service-Pakete aufweisen, effektiv differenzieren [Fl21c, Sc15].

Bei IT-Service-Katalogen werden in Wissenschaft und Praxis zwei grundlegende Sichten unterschieden, die externe Sicht für Nachfragende bzw. Kunden und die interne Sicht für die möglichst automatisierte Umsetzung und Erbringung durch die IT. Diese Sichten dienen unterschiedlichen Zwecken und erfordern daher spezifische Informationen [NJ15; NDB16, MA19; MS10; Ti20]: Bei der **externen Sicht** bzw. **Kundensicht** werden IT-Services, die sogenannten Business-Services, in der Sprache der Kunden beschrieben. Auf dieser Basis können Kunden unterschiedliche IT-Service-Request-Arten auslösen. Die Kunden können je nach Service-Angebot aus der Perspektive des Unternehmens extern oder intern (Mitarbeiter aus Fachabteilungen) sein. Um die Business-Services IT-seitig erbringen zu können, werden technische Services (**interne Sicht**) benötigt. Sie definieren Verantwortlichkeiten und gleichermaßen manuelle und digitale Aufgaben, wie die Installation, den Betrieb und die Wartung einer bestimmten Applikation. Während Business-Services mit Geschäftsprozessen der Kunden verlinkt sind, werden technische Services einerseits mit Business-Services und andererseits mit den für ihre Erbringung notwendigen Configuration-Items (CIs) wie Anwendungs- und Serverinstanz verknüpft. Die interne Sicht umfasst zudem die Definition von Genehmigungsprozessen und die zur Beschaffung und Entsorgung von IT-Komponenten erforderlichen Aktivitäten.

Beide Sichten ergeben zusammen eine durchgängige und transparente Service-Struktur, welche die Grundlage für zahlreiche ITSM-Prozesse darstellt. Zum einen lässt sich bei einem Ausfall einer konkreten Serverinstanz durch das Incident-Management effizient ermitteln, welche technischen Services und damit verbundene Business-Services bei den dazugehörigen Kunden betroffen sind. Zum anderen kann bei einer Störungsmeldung auf der Business-Service-Ebene die Ursache mithilfe der Nachverfolgung der Verknüpfungen bei der internen Sicht identifiziert werden [Ax19; NJ15; NDB16]. Untersuchungen zeigen allerdings, dass bei Weitem nicht allen Unternehmens- und IT-Verantwortlichen die Bedeutung beider Sichten und deren notwendige enge Verzahnung bewusst ist. Vielfach wird nur eine davon, vorrangig die Kundensicht, adressiert [Fl21c; MA19]. Dieser Umstand manifestiert sich auch darin, dass unter IT-Service oftmals ausschließlich der Business-Service-Teil gemeint ist und somit die Begriffe IT-Service und Business-Service synonym verwendet werden, ohne den internen Teil abzugrenzen oder explizit zu erwähnen. Die Etablierung der internen Sicht mit der Definition von technischen Ser-

vice- und CI-Arten erweist sich als besondere Herausforderung [Sc17a; SH18]. Es handelt sich jedoch nicht nur um eine begriffliche Unschärfe, sondern durch die partielle Erfassung der Leistung bleibt viel Potential von ITSM ungenutzt. In der Folge lässt sich für gewöhnlich keine hinreichende Standardisierung und Automatisierung von Prozessen bei der Umsetzung, der Erbringung sowie dem Monitoring von IT-Services erreichen. Außerdem wird eine transparente Kostenkalkulation und Preisfindung erschwert.

Dieser Beitrag adressiert mit einem Fokus auf die externe Sicht praktisch das Schaufenster der IT. Die Gründe dafür sind die größere Verbreitung in der Praxis und die Tatsache, dass oftmals zunächst auf Basis der Kundenanforderungen, einer Wettbewerbsanalyse und der gegenwärtigen IT-Landschaft Business-Services definiert werden. Erst danach wird deren Umsetzung im Detail geplant [HL18]. Sobald die Service-Katalog-Grundstruktur vorliegt, müssen die einzelnen IT-Services anhand geeigneter Attribute einheitlich und kundenzentriert beschrieben werden [Fl21c; MA22; MA19; NJ15]. Zur Festlegung und Befüllung der Attribute gibt es in der Literatur mehrere Ansätze, die aber nicht aufeinander abgestimmt sind [Fl21a; PZ19; RBK08]. Eine kritische Bewertung und die Konsolidierung stellen damit eine aktuelle Herausforderung dar [MA22; MA18].

3 Forschungsdesign

Zur Zielerreichung wurde ein zweistufiger Ansatz, bestehend aus einer strukturierten Literaturrecherche gefolgt von einer vergleichenden Analyse und Synthese, gewählt. Der erste Schritt bestand in der Identifizierung der bereits vorliegenden Ansätze zur Beschreibung von IT-Services. Hierfür wurde eine strukturierte Literaturrecherche gemäß den Richtlinien von Webster und Watson [WW02] durchgeführt. Knackstedt und Winkelmann [KW06] folgend wurden sowohl Datenbanken, die eine interdisziplinäre Suche ermöglichen, als auch Datenbanken mit einem Fokus auf einen bestimmten Themen- und Fachbereich selektiert: AIS eLibrary, EBSCO Host, Emerald Insight, Google Scholar und Science Direct. Als relevant wurden Beiträge eingestuft, die einen eigenen Ansatz zur Beschreibung von IT-Services aus Kundensicht vorschlagen, einen vorhandenen modifizieren oder mehrere existierende aufführen bzw. miteinander vergleichen. Auf Basis dieser Suchergebnisse erfolgte eine Vor- und Rückwärtssuche, wobei als Ergebnis letztlich 24 Publikationen ermittelt werden konnten (Anhang, Tabelle 2).

Im zweiten Schritt wurde die in Schritt eins generierte Literaturbasis den Richtlinien von Webster und Watson [WW02] entsprechend mit einem konzeptzentrierten Bottom-up-Ansatz untersucht. Alle in der Literatur genannten Attribute wurden dazu aufgelistet, bevor sie in der Konzeptmatrix (auf Anfrage bei den Autoren erhältlich) iterativ aggregierten Konzepten zugeordnet wurden. In diesem Aggregationsschritt wurden Synonyme und unterschiedliche Abstraktionsebenen harmonisiert. Es erfolgte keine Prüfung der Attribute auf Plausibilität. Die Attribute wurden nach ihrer Auftrittshäufigkeit sortiert, um dadurch erste Hinweise auf ihre Relevanz zu erhalten. Die Herausforderung bestand darin, Überschneidungen bei den resultierenden Attributen zu vermeiden und gleich-

zeitig ein Höchstmaß an Vollständigkeit zu erreichen. Die Erstellung der Attributliste wurde dadurch erschwert, dass Veröffentlichungen teils Attribute nur auflisten bzw. im Text erwähnen, ohne sie zu erläutern.

4 Ergebnisse

Die in den 24 Publikationen genannten 228 Attribute konnten zu 69 abgrenzbaren Attributen konsolidiert werden (Tabelle 1). Die Anzahl der vorgeschlagenen Attribute reichte von 3 bis 20 (Mittelwert: 9,5) und variiert somit erheblich. Die Beiträge mit den meisten Attributen sind die von Floerecke [Fl21c] (20), Mayerl et al. [ML05] (19) sowie Mendes und da Silva [MS10] (17). Lediglich vier Attribute (Kurzbeschreibung, Name, Preis und SLAs) sind Bestandteil einer Mehrheit der Studien. Zahlreiche Attribute werden nur ein einziges Mal (27) oder zweimal (13) aufgeführt, was aber nicht zwangsläufig bedeutet, dass diese weniger bedeutsam oder gar überflüssig sind. Im Durchschnitt kommt jedes Attribut 3,3-mal vor. Der Median beträgt 2.

Insgesamt weisen die vorhandenen Beschreibungsansätze, sowohl die Anzahl als auch die Ausprägung der einzelnen Attribute betreffend, deutliche Unterschiede auf. Ein gemeinsamer Kern an Attributen im engeren Sinn lässt sich aktuell nicht erkennen. Die am häufigsten genannten Attribute (n>5) in absteigender Reihenfolge sind *Kurzbeschreibung*, *Name*, *Preis*, *SLAs*, *Leistungsbestandteile (kurz)*, *Service-Katalog-Oberkategorie*, *IT-Service-Owner*, *ID*, *Leistungsreporting*, *Status* und *Kontakt.* Tabelle 1 zeigt die konsolidierte Attributliste. Enthalten sind die 69 IT-Service-Beschreibungsattribute in absteigender Auftrittshäufigkeit mitsamt einer kurzen Erläuterung. Zudem ist der Datentyp angegeben, der den Formalisierungsgrad und damit die Automatisierungsmöglichkeiten und die Wiederverwendbarkeit der Einträge ausdrückt. Hinzugefügt ist außerdem, inwieweit ein Attribut auch für die interne Sicht des IT-Service-Katalogs von Bedeutung sein könnte (niedrig, mittel, hoch). Die relativ häufig gesetzten Werte „hoch“ und „mittel“ spiegeln die enge Verzahnung der internen und externen Sicht wider. Bei den Kriterien „Datentyp“ und „Relevanz interne Sicht“ handelt es sich um eine auf der wissenschaftlichen und praktischen Erfahrung der Autoren basierende, erste Qualifizierung, die einer weiteren Diskussion bedarf.

Attribut	Kurzerläuterung	Datentyp	Relevanz interne Sicht
Kurzbeschreibung	Allgemeine Beschreibung des IT-Service in Kurzform	text	niedrig
Name	Bezeichnung des IT-Service aus Kundensicht	text	mittel
Preis	Preis des IT-Service (Aktivierungsgebühr und Einmal-/monatlicher Preis)	float	niedrig
SLAs	Darstellung der SLAs (Verfügbarkeit, Reaktionszeit des Anbieters, etc.)	enum	hoch
Leistungsbestandteile (kurz)	Auflistung der integrierten Leistungsbestandteile des IT-Service	text	hoch

Service-Katalog-Oberkategorie	Übergeordnete Strukturierungsebene im IT-Service-Katalog (Ebene 1)	enum	niedrig
IT-Service-Owner	Verantwortlicher für den IT-Service entlang des gesamten Lebenszyklus	text	hoch
ID	Eindeutige Identifikationsnummer des IT-Service im IT-Service-Katalog	int	niedrig
Leistungsreporting	Ergebnis und Qualität vergangener Service-Erbringungen (Soll vs. Ist)	text/media	hoch
Status	Aktuelle Lebenszyklusphase (in Produktion, in Ablösung, etc.) des IT-Service	enum	hoch
Kontakt	Allgemeine Kontaktdaten für am IT-Service interessierte Akteure	text	niedrig
Bezugsvoraussetzungen	Anforderungen technischer und prozessualer Art an den IT-Service-Bezieher	text	hoch
Organisationsbereich	Der IT-Service bereitstellende interne Organisationsbereich	text	hoch
Version	Version des IT-Service	float	mittel
Technische Informationen	Technische Fakten zum IT-Service (Rechenleistung, Speicherkapazität, etc.)	text	hoch
Exkludierte Leistungen	Vom IT-Service exkludierte Leistungen und Bestandteile	text	hoch
Kreis der Bestellberechtigten	Für den IT-Service-Bezug zugelassene Personen/Rollen	enum	mittel
Leistungsbestandteile (lang)	Ausführliche Beschreibung der integrierten Bestandteile des IT-Service	text	niedrig
Langbeschreibung	Allgemeine Beschreibung des IT-Service in Langform	text	niedrig
Stellvertreter	Stellvertreter für in der IT-Service-Beschreibung aufgeführter Rollen	text	niedrig
Technische Umsetzung	Grundlegende Angaben zur technischen Umsetzung des IT-Service	text	hoch
SLA-Parameter	Allgemeine Beschreibung der in den SLAs enthaltenen Parametern	text	niedrig
Kritikalität	Höhe des potenziellen Schadens bei Ausfall des IT-Service auf Kundenseite	text	hoch
Quality of Service (QoS)	Zugesicherte Güte des Kommunikationsdienstteils des IT-Service	enum	hoch
Gültigkeitsbereich	Für den IT-Service freigeschaltete Organisationseinheiten	enum	mittel
IT-Service-Nutzen	Darlegung des durch den IT-Service-Bezug entstehenden Nutzen für Kunden	text	niedrig
Eskalationsverfahren	Beschreibung des Problemlösungs- und Eskalationsprozesses	text	niedrig
Verfügbarkeit	Angabe zur aktuellen Verfügbarkeit bzw. des Lagerbestands des IT-Service	binär/enum	hoch
Beschreibungsverantwortlicher	Verantwortlicher und/oder Autor der IT-Service-Beschreibung	text	niedrig
Service-Katalog-Subkategorie	Strukturierungsebene unterhalb der Service-Katalog-Oberkategorie (Ebene 2)	enum	niedrig
Unterstützende IT-Services	Den Haupt-IT-Service unterstützende IT-Services (z.B. Backup-Dienste)	text	hoch
Business-Owner	Verantwortlicher für die Business-Service-Ebene des IT-Service	text	mittel
IT-Service-Manager	Strategischer Verantwortlicher des IT-Service	text	mittel
Hersteller	Herstellernamen der Service-Kernkomponenten (z.B. Applikation, Hardware)	text	mittel
Abbildungen	Bilder/Videos (grundlegende Vorstellung, Bedienung, etc.) zum IT-Service	media	niedrig
Wartung	Beschreibung des Prozesses der geplanten und ungeplanten Wartung	text	hoch
Optionale Leistungen	Innerhalb des IT-Service optional hinzubuchbare Leistungen	enum	hoch
Bezugsgröße	Bezugsgröße des IT-Service (z.B. User, Endgeräte)	enum	mittel
IT-Service-Varianten	Auflistung alternativer Varianten des IT-Service	enum	hoch
Zusatzinformationen	Zusatzinformationen jeglicher Art (z.B. für einen speziellen	text	niedrig

	Industriezweig)		
Bestellprozess	Beschreibung des IT-Service-Bestellprozesses aus Kundensicht	text	niedrig
IT-Service-Typ	Nennung eines übergeordneten IT-Service-Typs zur Klassifikation	enum	niedrig
Gewicht und Maße	Gewicht und Maße des IT-Service (bei Hardware)	float	niedrig
Vertragslaufzeit	Vertragslaufzeit des IT-Service-Bezugs	date	mittel
Lieferzeit	Durchschnittliche Dauer von der Bestellung bis zur IT-Service-Bereitstellung	date	hoch
Ähnliche IT-Services	Auflistung ähnlicher IT-Services	enum	hoch
Letzte Beschreibungsänderung	Datum der letzten Aktualisierung der IT-Service-Beschreibung	date	niedrig
Preiskategorie	Preisklassifizierung (z.B. günstig, teuer) des IT-Service (interne Kunden)	enum	niedrig
SLA-Varianten	Zur Verfügung stehende, wählbare SLA-Varianten des IT-Service	enum	hoch
Support	Beschreibung des Support-Angebots und -Prozesses	text	mittel
Abhängigkeiten	Zur Nutzung zusätzlich erforderliche IT-Services (z.B. Client für Anwendung)	enum	hoch
Kundenmitwirkungspflichten	Mitwirkungspflichten des Kunden bei der Service-Order und -Bereitstellung	text	hoch
Nutzungsbedingungen	Nutzungsbedingungen des IT-Service (z.B. maximale Umgebungstemperatur)	enum	hoch
Vertriebsweg	Möglichkeiten des IT-Service-Erwerbs (z.B. Telefon, Self-Service-Portal)	enum	niedrig
Strafen	Vereinbarte Vertragsstrafen bei Nichteinhaltung der SLAs auf Anbieterseite	float	hoch
Benötigte Genehmigung	Erforderliche Genehmigungen für die IT-Service-Bestellung (interne Kunden)	enum	mittel
Lieferprozess	Beschreibung des IT-Bereitstellungsprozesses	text	hoch
Qualitätszertifizierungen	Auflistung seitens des IT-Service erfüllter Zertifizierungen	enum	hoch
Quellen	Für die Service-Beschreibung verwendete, öffentlich zugängliche Quellen	text	niedrig
Kompatible IT-Services	Mit dem IT-Service kompatible IT-Services	enum	hoch
Vertragliche Details	Beschreibung vertraglicher Details	text	hoch
Tipps zur Kosteneinsparung	Hinweise zur kosteneffizienten Nutzung (primär interne Kunden)	text	niedrig
Geschäftsprozessunterstützung	Auflistung der vom IT-Service unterstützten Geschäftsprozesse	enum	niedrig
Geschäftsfähigkeitsabdeckung	Nennung der vom IT-Service abgedeckten Geschäftsfähigkeiten	enum	niedrig
Infrastrukturressourcen	Für den IT-Service notwendige Infrastrukturressourcen auf Anbieterseite	enum	hoch
IT-Service-Anbieter	Bei der IT-Service-Erbringung beteiligte Anbieter (z.B. interne IT + Cloud)	enum	hoch
Service-Request-Arten	Beschreibung der unterschiedlichen, angebotenen Service-Request-Arten	text	hoch
Benutzerhandbücher	Link/sonstiger Bezugsweg von Benutzerhandbüchern	text/link	niedrig
Vorgänger/Nachfolger	Auflistung von Vorgänger-/geplanten Nachfolger-IT-Services	enum	hoch

Tab. 1: Konsolidierter IT-Service-Attributkatalog

Bei der Erstellung der Konzeptmatrix, der Konsolidierung und Zuordnung der identifizierten Attribute, ist von einem gewissen Grad an Subjektivität auszugehen. Denn die Attribute zeigten teils einen unterschiedlichen Abstraktionsgrad, der gegebenenfalls eine

weitere Klärung erfordert. Hinzu kommt die Tatsache, dass manche Publikationen keine oder nur unzureichende Erklärungen der Attribute beinhalten. Die ITSM-Rahmenwerke waren nicht Gegenstand und deren Betrachtung inklusive des Vergleichs mit den Ergebnissen dieser Untersuchung steht noch aus. Die Analyse der 24 Publikationen nach den Kriterien Forschungsmethode, Veröffentlichungsjahr und -sprache sowie Publikationsart führte zu folgenden Ergebnissen:

Methode: Die 24 Quellen basieren auf Literaturrecherchen, Fallstudien (Experteninterviews, Workshops und Dokumentenanalysen) und Erfahrungen einzelner Autoren in der Praxis. In vielen Publikationen wurden mehrere Methoden angewandt. Bei zwei ([An05; SS16]) fehlt eine entsprechende Angabe. Die Untersuchungen bauen kaum aufeinander auf. Insgesamt sind die bisherigen Ergebnisse abstrakt und erscheinen unvollständig und damit wenig an den Bedürfnissen der Praxis ausgerichtet.

Veröffentlichungsjahr: Es konnten keine Beiträge vor 2005 ermittelt werden. Bemerkenswert ist, dass wenige aktuelle Studien vorliegen. So wurden zwischen 2020 und März 2022 lediglich vier Beiträge publiziert. Es scheint, als würde dieses Thema für die Mainstream-Forschung im Bereich der Informatik/Wirtschaftsinformatik wenig Attraktivität bieten. In den Jahren 2009 und 2010 erreichte das Interesse seinen bisherigen Höhepunkt (vier bzw. drei Veröffentlichungen). Dieser Peak könnte unter anderem auf die Veröffentlichung von ITIL v3 im Jahr 2007 mit der erstmaligen Einbindung des ITSM-Prozesses IT-Service-Katalog-Management zurückgeführt werden. Die beiden vor 2007 erschienenen Beiträge lassen den Schluss zu, dass das Service-Katalog-Konzept zwar bereits zuvor in der Literatur diskutiert, aber erst durch ITIL bekannter wurde.

Veröffentlichungssprache: 11 der 24 Publikationen (46 %) sind in deutscher Sprache verfasst. Der relativ hohe Anteil unterstreicht die große ITSM-Community innerhalb der DACH-Region. Beim itSMF beispielsweise sind über 1600 Fach- und Führungskräfte an der Steuerung der Entwicklung des ITSM und der Verbreitung der daraus entstehenden Erkenntnisse beteiligt.

Publikationsart: 42 % der Beiträge wurden in wissenschaftlichen Zeitschriften, 33 % in Konferenz- und Tagungsbänden veröffentlicht. Den Rest machen Bücher/Buchkapitel (13 %), Masterarbeiten, Whitepaper und Internetauftritte (je 4 %) aus. Die erkennbare Fragmentierung der Literatur wird durch die Analyse der Zitierungen bei Google Scholar bestätigt (Mittelwert: 6,5). Mit Ausnahme von [NJ15] wurde keiner der Journal-Beiträge in einer einflussreichen und hochrangigen wissenschaftlichen Zeitschrift gemäß dem VHB-JOURQUAL3-Ranking publiziert.

5 Kritische Reflexion und Ausblick

Vorliegende Studie ermittelte, analysierte und synthetisierte die fragmentierte wissenschaftliche Literatur zur Beschreibung von IT-Services aus Kundensicht. Zentrales Ergebnis ist ein aus 69 Attributen bestehender Attributkatalog. Die vorhandenen Beschrei-

bungsansätze weisen, sowohl die Anzahl als auch die Ausprägung der einzelnen Attribute betreffend, deutliche Unterschiede auf. Ein gemeinsamer Kern an Attributen im engeren Sinn ist nicht zu erkennen. Zudem ist festzuhalten, dass vergleichsweise wenige Forschungsbeiträge zu diesem Thema vorliegen. Die Beiträge wurden überwiegend nicht in führenden wissenschaftlichen Zeitschriften und Konferenzen publiziert. Was die angewandten Methoden anbelangt, macht die Erhebung von Praxiserfahrungen durch Experteninterviews, Workshops und Dokumentenanalysen den Hauptteil aus. Dennoch sind die Forschungsergebnisse abstrakt geblieben und damit wenig auf die Bedürfnisse der Praxis ausgerichtet. Dieser Umstand ist erstaunlich, da die Beschreibung von IT-Services für die Praxis branchenübergreifend ein wichtiges und zentrales Anliegen darstellt.

Nicht jedes der Attribute ist für Kunden von gleicher Bedeutung oder auch aus rechtlicher Sicht notwendig. Für eine weiterführende Evaluation könnte sich eine Online-Umfrage mit IT-Verantwortlichen und ITSM-Experten, aber auch mit internen und externen Kundenvertretern unterschiedlicher Branchen eignen (Referenz für ein mögliches Forschungsdesign: [FRL20]). Dabei wird angeregt, die Möglichkeit einer Typisierung oder Bildung von disjunkten Attributmengen zu untersuchen. Kriterien einer Attributmengenaufteilung könnten mitunter Nachfolgende sein: (1) Service-Katalog als Dokument vs. Bestandteil eines Self-Service-Portals, (2) Unternehmensgröße und Branche (branchenspezifische Regularien), (3) Interne und/oder externe Kunden als Zielgruppe (Monopol vs. Konkurrenzsituation), (4) IT-Service-Art (Hardware vs. Software / Einzelleistung vs. Service-Paket) und (5) Service-Request-Arten. Denn die Vermutung liegt nahe, dass es zwar keinen allgemeingültigen Referenzrahmen geben dürfte, aber abhängig von verschiedenen Faktoren eine gewisse Standardisierung möglich sein könnte.

Angesichts der Ergebnisse dieser Metastudie ist nicht davon auszugehen, dass bereits sämtliche relevanten Attribute bekannt sind und selbst der konsolidierte Attributkatalog als Grundlage für vollumfängliche IT-Service-Beschreibungen ausreicht. Die Identifikation weiterer Attribute könnte in die vorgeschlagenen, weiterführenden Studien einbezogen oder in separaten Untersuchungen ermittelt werden. Dazu könnten existierende Service-Kataloge von Unternehmen ausgewertet und damit zusätzliche Attribute abgeleitet werden. Vielversprechend könnte es zudem sein, Anbieter und Dienstleister von ITSM-Software-Suiten und Self-Service-Portalen zu befragen bzw. deren Tools zu analysieren. Nicht vergessen werden darf eine entsprechende, systematische Betrachtung der führenden ITSM-Rahmenwerke. Dennoch kann der geschaffene Attributkatalog als nützliche Referenz für die Erstellung bzw. Überarbeitung von IT-Service-Katalogen dienen. Ergänzend ist anzumerken, dass es nicht zielführend ist, präventiv möglichst viele Attribute einzubauen, denn sie blähen die IT-Service-Beschreibungen für Kunden gegebenenfalls nur unnötig auf und führen zu einem Mehr an Pflegeaufwand beim Anbieter.

Wenn innerhalb eines Unternehmens eine für geeignet erachtete Attributliste definiert worden ist, sollte als nächster Schritt für jedes der Attribute der Service-Beschreibung ein Template spezifiziert werden, um alle im IT-Service-Katalog enthaltenen IT-Services strukturiert mit Inhalt zu befüllen. Dabei geht es unter anderem um Verbal- oder Nominalstil, ganze Sätze oder Stichpunkte, Mindestlängen und Format von Datums-, Zeit- und

Preisangaben. Auch Aufzählungs- bzw. Auswahlfelder wie verschiedene Service-Level-Stufen oder Statusausprägungen sollten möglichst in der Sprache der Kunden gestaltet sein. Templates sind vor allem deshalb unverzichtbar, da in der Regel eine Vielzahl von Mitarbeitern an den Service-Beschreibungen arbeiten. Im Zusammenhang mit der Attributfestlegung empfiehlt es sich, Attribute nach Themenbereichen zu clustern (z.B. Basisinformationen, Verantwortlichkeiten sowie Preis- und Vertragsinformationen), um dadurch die Beschreibungen für Kunden übersichtlich zu halten. Hierzu bietet die Forschung mit [ML05] zumindest einen ersten rudimentären Vorschlag.

Getrennt vom IT-Service-Katalog im engeren Sinn zu betrachten ist, welche Informationen Kunden bei den unterschiedlichen Service-Request-Arten und IT-Services zur Verfügung stellen müssen. Dazu zählen beispielsweise Kunden-ID, Name, Büronummer, Kostenstelle oder Nummer der LAN-Buchse. Im Optimalfall sind die notwendigen Informationen bereits im System vorhanden – bei Eingabe des Namens in der ITSM-Software-Suite durch den IT-Service-Mitarbeiter bzw. beim erfolgreichen Login des Kunden im Self-Service-Portal. Dadurch werden der Eingabeaufwand für Kunden und Anbieter, aber auch die Gefahr von Falscheingaben bei der manuellen Erfassung reduziert. Der Eingang eines Service-Requests sollte einen Workflow in der ITSM-Software-Suite auslösen. Ein solcher Workflow legt fest, welche Aufgaben einzelne IT-Systeme und Personen manuell bzw. digital in welcher Reihenfolge abarbeiten müssen. Danach übernimmt der IT-Betrieb die dauerhafte IT-Service-Bereitstellung.

Um die Geschäftsführung und IT-Verantwortliche bei der IT-Service-Katalog-Erstellung zu unterstützen, bedarf es für jede der folgenden drei Komponenten einer praktikablen Lösung: Erstens wird ein **Vorgehensmodell** benötigt, welches vorgibt, welche grundlegenden Aktivitäten in welchen Phasen des Gestaltungsprozesses stattfinden müssen (z.B. [MB09; SS16]). Zweitens ist ein **Datenmodell** erforderlich, das festlegt, welche Artefakte in jeder einzelnen Phase produziert werden, welche Beziehungen zwischen diesen bestehen und welche Daten diese enthalten müssen. Hier sind die erwähnten Templates von zentraler Bedeutung (z.B. [NJ15; NDB16]). Drittens ist ein **Methodenbaukasten** nötig, der spezifische Methoden und Techniken umfasst, die der Generierung der einzelnen Artefakte innerhalb des Vorgehensmodells dienen (z.B. [KK09; MA19]). Die drei Komponenten müssen in eine adäquate IT-Unterstützung eingebettet werden. Die bereits vorliegenden Ansätze müssen geeignet zusammengeführt und erweitert werden. Diese Metastudie hat dies für einen Teilbereich des Datenmodells vorgenommen.

Eine weitere, in der Praxis zunehmend an Bedeutung gewinnende, aber unzureichend erforschte Fragestellung ist, inwiefern sich die gesamthafte IT-Service-Struktur (die Basis des ITSM) mit den Informationen des Enterprise-Architecture-Managements (EAM) etwa hinsichtlich Anwendungen, technischen Bausteinen, CIs und Geschäftsprozessen konzeptuell und systemisch geeignet integrieren lässt. Tool-Hersteller (z.B. LeanIX und ServiceNow) haben diese Kundenanforderung bereits erkannt und bieten erste Integrationskonzepte- und Lösungen an, um eindeutige Masterschaften und Beziehungen zu definieren und Redundanzen zu vermeiden.

Abschließend lässt sich festhalten, dass sich die Forschung zwar mit der Frage „Anhand welcher Attribute sollten IT-Services aus Kundensicht beschrieben werden?“ bereits befasst hat, jedoch noch keine brauchbare Antwort für die Praxis vorliegt. Es existieren weder hinreichend praxistaugliche Vorlagen für die IT-Service-Grundstruktur noch für die Beschreibung einzelner IT-Services. Mit dem konsolidierten Attributkatalog wurde versucht, einen Beitrag zur Verbesserung zu leisten. Neben den bereits im Text erwähnten offenen Fragen ist ein wichtiger nächster Schritt die komplementäre Beschreibung der internen Sicht von IT-Service-Katalogen. Im Unterschied zur Kundensicht ist deren unternehmensseitige Abdeckung noch kaum untersucht. Grundsätzlich gilt aber, nur wenn beide Sichten des IT-Service-Katalogs abgedeckt und geeignet miteinander verknüpft sind, kann das gesamte Potential bei den ITSM-Prozessen ausgeschöpft werden.

Literatur

[An05] Anders, T.: Development of a Generic IT Service Catalog as Pre-Arrangement for Service Level Agreements. International Conference on Emerging Technologies and Factory Automation (ETFA), Catania, Italien, 2005.

[Ax19] Axelos: ITIL Foundation: ITIL 4 Edition. TSO, Norwich, 2019.

[BL19] Bause, M.; Landis, M.: Wie sericezentriert ist Ihre IT wirklich: Ergebnisse einer Online-Umfrage in der DACH-Region. USU GmbH, Möglingen, 2019.

[BZ14] Beims, M.; Ziegenbein, M.: IT-Service-Management in der Praxis mit ITIL. Carl Hanser, München, 2014.

[Ca16] Casson, D.: Why 80 Percent of Service Catalog Projects Fail. Evergreen, USA, 2016.

[Fl22] Floerecke, S.: Self-Service-Portale mit Schwerpunkt Service-Request-Management: Analyse der Eignung existierender Software-Tool-Kategorien. HMD – Praxis der Wirtschaftsinformatik, 59(2), S. 683-693, 2022.

[Fl21a] Floerecke, S.: Self-Service-Portale für das IT-Service-Request-Management: Identifikation relevanter Use-Cases. HMD – Praxis der Wirtschaftsinformatik, 2021. DOI: 10.1365/s40702-021-00804-1

[Fl21b] Floerecke, S.: Service-Request-Typen und ITSM-Prozesse richtig abdecken. Digital Business Cloud, 1/2021, S. 34-35, 2021.

[Fl21c] Floerecke, S.: Best-Practices für die Gestaltung von IT-Service-Katalogen und den Einsatz von Self-Service-Portalen. HMD – Praxis der Wirtschaftsinformatik, 58(4), S. 1508-1520, 2021.

[FRL20] Floerecke, S.; Röck, F. F.; Lehner, F.: Success-Driving Business Modell Characteristics of IaaS Providers – Towards a Quantitative Evaluation. International Journal on Cloud Computing Services and Architecture, 10(4), S. 1-10, 2020.

[Ga22] Gartner (2022): Gartner Glossary: IT-Services. Gartner, Stamford, USA, 2022.

[HL18] Hellinger, B.; Landis, M.: Entwicklung einer servicezentrierten IT-Organisation. USU GmbH, Möglingen, 2018.

[ID21] IDG Research Services: Studie IT Service Management 2021. IDG Research Services, München, 2021.

[IT07] IT Service Management Forum Deutschland: ITIL V3-Glossar. Arbeitskreis Publikation – ITIL Version 3 Translation Project. Frankfurt am Main, 2007.

[KB17] Kasulke S.; Bensch, J.: Zero Outage: Kompromisslose Qualität in der IT im Zeitalter der Digitalisierung. Springer Gabler, Wiesbaden, 2017.

[KW06] Knackstedt, R.; Winkelmann, A.: Online-Literaturdatenbanken im Bereich der Wirtschaftsinformatik: Bereitstellung wissenschaftlicher Literatur und Analyse von Interaktionen der Wissensteilung. Wirtschaftsinformatik, 48(1), S. 47-59, 2006.

[KK09] Kohlborn, T.; Korthaus, A.; Chan, T.; Rosemann, M.: Identification and Analysis of Business and Software Services – A Consolidated Approach. IEEE Transactions on Services Computing, 2(1), S. 50-64, 2009.

[MA22] Macias, C. M.; Aguilar-Alonso, I.: Assessment of Contributions of the Methodology for the Construction of a Catalog of Information Technology Services Aimed at Public Entities. International Journal of Service Science, Management, Engineering, and Technology, 13(1), S. 1-28, 2022.

[MA19] Macias, C. M.; Aguilar-Alonso, I.: Proposal for the Identification of Information Technology Services in Public Organizations. Symmetry, 11, S. 1269-1301, 2019.

[MA18] Macias, C. M.; Aguilar-Alonso, I.: Review of Proposals for the Construction and Management of the Catalog of Information Technology Services. IEEE Access, 6, S. 45335-45346, 2018.

[ML05] Mayerl, C.; Link, S.; Racke, M.; Popescu, S.; Vogel, T.; Mehl, O.; Abeck, S.: Methode für das Design von SLA-fähigen IT-Services. In: Kommunikation in Verteilten Systemen (KiVS) (S. 271-282). Springer, Berlin, Heidelberg, 2005.

[MB09] Mazvimavi, V.; Benyon, R. V.: A Theoretical Model for Developing an IT Service Catalogue. International Conference on Information Management (Conf-IRM 2009), Al Ain, United Arab Emirates, 2009.

[MS10] Mendes, C.; da Silva, M. M.: Implementing the Service Catalogue Management. 7th International Conference on the Quality of Information and Communications Technology, Porto, Portugal, 2010.

[NDB16] Nord, F.; Dörbecker, R; Böhmann, T.: Structure, Content and Use of IT Service Catalogs – Empirical Analysis and Development of a Maturity Model. 49. Hawaii International Conference on System Sciences, Manoa, Hawaii, 2016.

[NJ15] Nissen, V.; Jung, D.; Petsch, M.; Präg, C. P.: Recommendations for a General IT Service Catalogue Structure. Enterprise Modelling and Information Systems Architectures (EMISAJ), 10(1), S. 89-107, 2015.

[PZ19] Pröhl, T.; Zarnekow, R.: Agilität bei der Einführung von IT-Servicemanagement: Lösung klassischer Herausforderungen mit agilen Methoden. HMD – Praxis der Wirtschaftsinformatik, 56(2), S. 301-317, 2019.

[RR18] Reiss, M.; Reiss, G.: Praxisbuch IT-Dokumentation: Vom Betriebshandbuch bis zum Dokumentationsmanagement – die Dokumentation im Griff. 3. Auflage, Carl Hanser,

München, 2018.

[RBK08] Rudolph, S.; Böhmann, T.; Krcmar, H.: Struktur von IT-Servicekatalogen: Ein praxisorientierter Gestaltungsvorschlag für die Dokumentation des IT-Leistungsangebots. Multikonferenz Wirtschaftsinformatik, München, Deutschland, 2008.

[Sc17a] Scholderer, R.: IT-Servicekatalog: Services in der IT professionell designen und erfolgreich implementieren. dpunkt, Heidelberg, 2017.

[Sc17b] Schopf, V.: Vom IT-Shop zum Businesstreiber – Servicekataloge im ITSM. itSMF, 41, S. 38-43, 2017.

[SH18] Schorr, F.; Hvam, L.: The Use of Design-Science to Define Information Content Requirements for IT Service Catalogs. IEEE International Conference on Industrial Engineering and Engineering Management, Bangkok, Thailand, 2018.

[Sc15] Schulz, C.: IT-Katalogmanagement: Steuerungsinstrument für den internen Hard- und Softwarebestand. OBJEKTspektrum, 3/2015, S. 68-72, 2015.

[SS16] Sembiring, M.; Surendro, K.: Service Catalogue Implementation Model. 4. International Conference on Information and Communication Technology, Neu-Delhi, Indien, 2016.

[Ti22] Tiemeyer, E.: Handbuch IT-Management: Konzepte, Methoden, Lösungen und Arbeitshilfen für die Praxis. Carl Hanser, München, 2022.

[WTT14] Walterbusch, M.; Truh, S.; Teuteberg, F.: Hybride Wertschöpfung durch Cloud Computing. In Dienstleistungsmodellierung 2014 (S. 155-174). Springer, Wiesbaden, 2014.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. Management Information Systems Quarterly, 26(2), S. 13-23, 2002.

Anhang

ID	Autor	Jahr	Titel	Forschungsmethode	Publikationsart	Sprache
[1]	Anders, T.	2005	Development of a Generic IT Service Catalog as Pre-Arrangement for Service Level Agreements	Nicht genannt	Konferenz	Englisch
[2]	Floerecke, S.	2021	Best-Practices für die Gestaltung von IT-Service-Katalogen und den Einsatz von Self-Service-Portalen	Literaturstudie und Einzellfall-studie mit Experteninterviews	Journal	Deutsch
[3]	Floerecke, S.	2021	Self-Service-Portale für das IT-Service-Request-Management: Identifikation relevanter Use-Cases	Literaturstudie und Einzellfall-studie mit Experteninterviews	Journal	Deutsch
[4]	Gerber, N. et al.	2019	Referenzkatalog der ICT Services im Gesundheitswesen: Modell für ICT Service Management, Controlling und Benchmarking	Literaturstudie und Erfahrungen aus der Praxis	Journal	Deutsch
[5]	Gödde, M.; Bujotzek, J.	2009	IT-Servicekatalog und Service-Pricing	Erfahrungen aus der Praxis	Journal	Deutsch
[6]	Hellinger, B.; Landis, M.	2018	Entwicklung einer servicezentrierten IT-Organisation – Aufbau eines Servicekatalogs	Erfahrungen aus der Praxis	Whitepaper	Deutsch
[7]	Horvat, N.; Matija; J.; Mekovec, R.	2013	Specifying Service Catalog – A Case Study	Literaturstudie und Einzell-fallstudie zur Evaluation	Konferenz	Englisch
[8]	Macias, C. M.; Aguilar-Alonso, I.	2022	Assessment of Contributions of the Methodology for the Construction of a Catalog of Information Technology Services Aimed at Public Entities	Umfrage in der Praxis	Journal	Englisch
[9]	Mayerl, C. et al.	2005	Methode für das Design von SLA-fähigen IT-Services	Mulitple Fallstudie	Buchkapitel	Deutsch
[10]	Mazvimavi, V.; Benyon, R. V.	2009	A Theoretical Model for Developing an IT Service Catalogue	Literaturstudie	Konferenz	Englisch
[11]	Mendes, C.; Silva, M. M.	2010	Implementing the Service Catalogue Management	Literaturstudie und Einzell-fallstudie zur Evaluation	Journal	Englisch
[12]	Nissen, V. et al.	2015	Recommendations for a General IT Service Catalogue Structure	Literaturstudie und Experteninterviews	Journal	Englisch
[13]	Nord, F.; Dörbecker, R.; Böhmann, T.	2016	Structure, Content and Use of IT Service Catalogs – Empirical Analysis and Development of a Maturity Model	Literaturstudie und Multiple Fallstudie mit Externinterviews und Dokumentenanlyse	Konferenz	Englisch
[14]	Pfitzinger, B.; Bley, H.; Jestädt, T.	2010	Service Catalogue and Service Sourcing	Einzelfallstudie	Journal	Englisch
[15]	PLAIN IT AG	2021	IT Service Katalog – Praxishilfe für die Definition eines Katalogs	Erfahrungen aus der Praxis	Ausführlicher Webauftritt	Deutsch
[16]	Rabbi, M. F.	2009	An IT Service Taxonomy for Elaborating IT Service Catalog	Literaturstudie und Internetrecherche	Masterarbeit	Englisch
[17]	Rudolph, S.	2009	Modell zur servicebasierten Planung und Steuerung der IT-Infrastruktur im Mittelstand	Literaturstudie und Experninterviews	Buch	Deutsch
[18]	Rudolph, S.; Böhmann, T.; Krcmar, H.	2008	Struktur von IT-Servicekatalogen: Ein praxisorientierter Gestaltungsvorschlag für die Dokumentation des IT-Leistungsangebots	Workshops und Experteninterviews	Konferenz	Deutsch
[19]	Sembiring, M.; Surendro, K.	2016	Service Catalogue Implementation Model	Nicht genannt	Konferenz	Englisch
[20]	Scholderer, R.	2017	IT-Servicekatalog: Services in der IT professionell designen und erfolgreich implementieren	Literaturstudie und Erfahrungen aus der Praxis	Buch	Deutsch
[21]	Schorr, F.; Hvam, L.	2018	The Use of Design-Science to Define Information Content Requirements for IT Service Catalogs	Literaturstudie und Fallstudie, inklusive Evaluation	Konferenz	Englisch
[22]	Schulz, S.	2015	IT-Katalogmanagement: Steuerungsinstrument für den internen Hard-und Softwarebestand	Literaturstudie und Einzellfall-studie mit Interviews, Workshops und Dokumentenanalyse	Journal	Deutsch
[23]	Taconi, L. H.; Barros, R. M.; Zarpelão, B. B.	2014	GAIA Service Catalog: A Framework for the Construction of IT Service Catalogs	Literaturstudie und Fallstudie, inklusive Evaluation	Journal	Englisch
[24]	Xu, D. et al.	2010	ICT Service Catalogue Representation Method and Application	Literaturstudie und prototypische Implementierung	Konferenz	Englisch

Tab. 2: Ermittelte Literaturbasis

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1629

Literatur-Review: State of the Art des Wertbeitrags der IT zum Unternehmenserfolg

Gian-Luca Wagner[1], Prof. Dr. Helmut Beckmann[2]

Abstract: Der Wertbeitrag der IT stellt ein kontrovers diskutiertes und diffuses Thema dar. Hierbei existieren viele Auffassungen was unter dem Wertbeitrag der IT verstanden wird und wie sich dieser definiert. Weiter ist nicht genau bekannt, wie sich der Wertbeitrag der IT auf Unternehmensebene äußert und so messbar gemacht werden kann. Diese Arbeit untersucht mithilfe einer systematischen Literaturanalyse den State of the Art des Wertbeitrags der IT zum Unternehmenserfolg und schafft hierbei ein Verständnis, was konkret unter dem Wertbeitrag der IT verstanden wird, wie sich dieser im Unternehmenskontext äußert und welche Ansätze existieren, diesen festzustellen.

Keywords: Wertbeitrag der IT, State of the Art, Unternehmen, Verständnis des Wertbeitrags, Ansätze zur Messbarkeit

1 Einleitung

Der Wertbeitrag der IT stellt in der Literatur ein lang diskutiertes Thema dar. Erkenntnisse aus dem empirischen Bereich gehen hauptsächlich auf die anglo-amerikanische Forschung zurück. Diese ordnen den Begriff des IT-Wertbeitrags unter den Begriffen „Business value of IT“ oder „IT-value proposition“ ein. In der Praxis werden hingegen Begriffe wie Effizienz, Leistung oder Produktivität genutzt, um den IT-Wertbeitrag zu beschreiben. Der IT-Wertbeitrag wird von Praktikern und Wissenschaftlern jedoch kritisch gesehen, da keine konkrete Antwort auf die Frage existiert, ob die Funktion der IT überhaupt einen Beitrag zur Erreichung strategischer Unternehmensziele leistet [Wi14]. Es existieren dennoch Ansätze den Beitrag der IT messbar zu machen, sogar Anhand von Erfolgsfaktoren. Ein wesentlicher Bestandteil der IT im Unternehmen stellt hierbei das Informationsmanagement dar, welches im Kontext der Wirtschaftsinformatik den optimalen Einsatz des Produktionsfaktors Information gewährleistet und als Ziel, die Informationen zum gewünschten Zeitpunkt in verarbeitbarer Konzentration bereitzustellen verfolgt. Dies definiert einen wesentlichen Erfolgsfaktor im Unternehmen [BDK21].

Problemstellung & Zielsetzung - Die Thematik, welchen Wertbeitrag die IT und ihre Funktionen am Unternehmenserfolg leistet ist laut [Wi14] diffus und wird in der Literatur

[1] Hochschule Heilbronn, Wirtschaft und Verkehr, Max-Planck-Str. 39, 74081 Heilbronn, gwagner@stud.hs-heilbronn.de
[2] Hochschule Heilbronn, Wirtschaft und Verkehr, Max-Planck-Str. 39, 74081 Heilbronn, helmut.beckmann@hs-heilbronn.de

doi:10.18420/inf2022_140

kontrovers diskutiert. Dennoch werden Teilgebiete der IT, bspw. das Informationsmanagement, wie [BDK21] beschreibt, als definierter Erfolgsfaktor gesehen, welcher das Ziel besitzt, einen optimierten Einsatz von Informationen im Unternehmen zu gewährleisten und so auch zum Erfolg des Unternehmens beiträgt. Es existieren somit konkrete Anhaltspunkte, dass die IT zum Wert des Unternehmenserfolges beiträgt. [TND14] bestätigen dies mit der Aussage, dass zahlreiche quantitative sowie qualitative Modelle und Vorgehensweisen zur Ermittlung des Wertbeitrags der IT existieren, woraus sich die Tatsache ergibt, dass die IT einen Beitrag zum Wert des Unternehmenserfolges leistet. Weiter beschreiben sie dennoch das Problem, dass diese Modelle keine umfassende und verlässliche Aussage über den Erfolg oder Misserfolg der IT im Unternehmen treffen [TND14]. Hierdurch ergibt sich das konkrete Problem, was unter dem Wertbeitrag der IT zum Unternehmenserfolg verstanden wird und welche Ansätze bestehen, diesen festzustellen. Diese wurde zum Anlass genommen, um den allgemeinen State-of-the-Art des Wertbeitrags der IT zum Unternehmenserfolg darzustellen. Aus der oben beschriebenen Problemstellung ergeben sich die folgenden zentralen Forschungsfragen des Beitrags: Q1 *Was wird unter dem Wertbeitrag der IT zum Unternehmenserfolg verstanden und wie äußert sich dieser?* Q2 *Welche Ansätze bestehen, den Wertbeitrag der IT zum Unternehmenserfolg festzustellen?* Ziel dieser Arbeit ist es daher den aktuellen Stand des Wertbeitrag der IT zum Unternehmenserfolg abzubilden. Zudem wird ein Verständnis des Wertbeitrags der IT zum Unternehmenserfolg geschaffen und die Äußerung dessen aufgezeigt. Weiter werden Ansätze darüber geliefert, wie der Wertbeitrag der IT zum Unternehmenserfolg festgestellt werden kann. Ein Nebeneffekt dieser Arbeit ist es, die Thematik des Wertbeitrags der IT im Kontext des Unternehmenserfolges zu aktualisieren und einen Anreiz für zukünftige Forschungen in diesem Bereich zu schaffen.

Methodik und Aufbau der Arbeit - Die Arbeit umfasst, nach einer kurzen Einleitung in das Thema und Problem- sowie Zielsetzung, eine ausführliche Darstellung der Ergebnisse mit anschließender kritischer Betrachtung der Arbeit und Ausblick auf weitere Forschung. Die Arbeit wird als Literature-Review nach dem Framework von [Br09] aufgebaut, welches vom Framework von [WW02] zur Erfassung der Ergebnisse ergänzt wird. Die Datenbanken und Konferenzen für Literaturquellen belaufen sich bei dieser Arbeit auf *Google Scholar*, *IEEE Xplore Digital Library*, *ScienceDirect*, *Springer Link*, *European Conference on Information Systems, International Conference on Information Systems* und *Gesellschaft für Informatik*.

2 Ergebnisse

Die Kriterien zur Durchführung wurden in das Hauptthema des Verständnisses des Wertbeitrags der IT zum Unternehmenserfolg eingeteilt, welches weiter in die Subthemen Definition (Wird der Begriff des Wertbeitrags der IT definiert?), Erläuterung (Besteht eine ausführliche Erläuterung des Wertbeitrags der IT?) und in welchem Kontext der Wertbeitrag der IT zum Unternehmenserfolg vorkommt, gegliedert wurde. Die Bereiche

basieren auf dem Beitrag von [TND14]. Das zweite Hauptkriterium wertet aus, ob in dem Beitrag Ansätze zur Messung bzw. Feststellung des Wertbeitrags der IT zum Unternehmenserfolg bestehen oder nicht.

Quelle	Verständnis des Wertbeitrags der IT							Ansätze zur Feststellung des Wertbeitrags
			Kontext					
	Definition	Erläuterung	(Business-)IT-Aglinment	IT-Governance	IT-Management	IT-Controlling	Allgemeiner Unternehmens-kontext	Methoden zur Messbarkeit/ Ermittlung
[Sp14]	x	x			x		x	x
[GM14]		x				x		x
[Af15]	x	x					x	x
[Ba15]	x	x	x				x	x
[Kil5a] [Kil5b]		x	x					x
[Sa15a]	x	x					x	x
[Sa15b]							x	x
[Kr16a]	x	x	x	x				
[Kr16b]							x	x
[PJ16]	x	x					x	x
[Fr17]	x	x	x					
[Kil7a] [Kil7b]	x					x		x
[SSW19]	x	x					x	x
[PKS20]	x				x			
[SG20]							x	x
[TMD20]	x						x	x
[SWW21]	x	x					x	x
Summe: 17	**12**	**11**	**4**	**1**	**2**	**2**	**11**	**14**
Prozent: 100 %	**70,59%**	**64,71%**	**23,53%**	**5,88%**	**11,76%**	**11,76%**	**64,71%**	**82,35%**

Abb. 1: Konzeptmatrix nach [WW02]

2.1 Verständnis des Wertbeitrags der IT

Definition

2014 - Eine der ersten Definitionen liefern [Sp14] im Bereich des allgemeinen Unternehmenskontextes, welche sich auf die Definition der RWTH Aachen stützen, in dem der Wertbeitrag der IT den Beitrag der IT zur Wettbewerbsfähigkeit eines Unternehmens darstellt [Sp14].

2015 - [Af15] definieren den Begriff des IT-Wertbeitrags ebenfalls im allgemeinen Unternehmenskontext, in dem der Wertbeitrag der IT als die Auswirkung der IT auf die Unternehmensleitung, sowohl auf der Zwischenprozessebene als auch auf der Ebene der gesamten Organisation definiert wird. Er umfasst laut [Af15] die Auswirkung auf die Effizienz als auch die Auswirkung auf die Wettbewerbsfähigkeit [Af15]. Eine etwas konkretere Definition liefert [Ba15] in Bezug auf den Wertbeitrag der IT, dieser wird zwar allgemein definiert, jedoch im Kontext des (Business-) IT-Alignment. Der IT-Wertbeitrag wird als das Ausmaß der Erreichung von Unternehmenszielen aufgefasst, welches durch den Einsatz von IT-Ressourcen im IT-Leistungserstellungsprozess erreicht wird (Teilgebiet des (Business-) IT-Alignment) [Ba15]. [Sa15a] definiert den Begriff des Wertbeitrags der IT, inwieweit die IT verantwortlich ist sowie sich in der Lage befindet, zur Leistungsfähigkeit eines Unternehmens am Markt beizutragen [Sa15a].

2016 - Im Beitrag von [Kr16b] wird der IT-Wertbeitrag als der Anteil des positiven Einflusses bzw. Mehrwert des IT-Einsatzes auf den Unternehmenserfolg definiert [Kr16a] [Kr16b].

2017 – [Fr17] definiert den Wertbeitrag der IT als komplementären Effekt der IT auf Prozesse, Menschen, strategische Steuerungsfunktionen, Kultur oder Wissensvermittlung [Fr17]. Der Wertbeitrag der IT wird von [Ki17b] sehr kurz und prägnant definiert als unternehmerische Wertsteigerung durch IT [Ki17a] [Ki17b].

2019 – Ebenfalls eine sehr kurze und prägnante Definition liefern [SSW19], in dem der IT-Wertbeitrag als monetäre Bewertung, der in einer Wirkung erfassten Beobachtung, definiert wird. Dies im Kontext der IT.

2020 – [PKS20] definieren den Begriff des Wertbeitrags der IT nahezu identisch mit dem von [Af15] wobei zusätzlich definiert wird, dass der IT-Wertbeitrag im Wesentlichen auf die Produktivitätssteigerung, Kostensenkung, Wettbewerbsvorteile, Bestandsreduzierung und anderen Messgrößen der Leistung einhergeht. Weiter definieren [SG20], dass der IT-Wertbeitrag eng mit IT-Projekten verknüpft ist und hierbei Kostensenkung, Umsatzsteigerung und Umsatzsicherung umfasst [SG20].

2021 – Die Definition von [SWW21] und [Se21] sind exakt dieselben, da die Arbeiten auf derselben Grundlage basieren. Der Wertbeitrag der IT wird hierbei als der Einfluss der Informationstechnologie (IT) auf die Unternehmensleistung definiert [SWW21] [Se21].

Zusammenfassend kann gesagt werden, dass sich der Kern des Wertbeitrags der IT auf den Mehrwert bezieht, den die IT zum Unternehmenserfolg leistet. Hierbei spielen Faktoren wie IT-Ressourcen eine wichtige Rolle, welche sich auf die Effizienz und die Wettbewerbsfähigkeit eines Unternehmens auswirken. Eng gekoppelt ist der IT-Wertbeitrag zudem an die Begriffe Produktivitätssteigerung und Kostensenkung, die meist mit einer Leistung der IT einhergehen. Deutlich wird auch, je später das Jahr der Publikation (nach 2021 aufsteigend), desto kürzer und prägnanter die Definitionen werden, wodurch interpretiert werden kann, dass der Wertbeitrag der IT im Kontext der Unternehmen angekommen ist, da ein gewisses Verständnis in der Thematik bereits aufgebaut wurde und keine detaillierten Erklärungen mehr zur Beschreibung nötig sind. Durchweg wird jedoch bis heute (2022) der Begriff des Wertbeitrag der IT definiert, was dafürspricht, dass dieser eine gewisse Relevanz besitzt und immer weiter spezifiziert und verfeinert wird.

Erläuterung und Äußerung

2014 - Im Beitrag von [Sp14] findet zur Definition auch eine kurze Erläuterung des Verständnisses des IT-Wertbeitrags statt. Diese wird hierbei als das richtige Handeln der IT im Unternehmenskontext gesehen, welches sich auf die Faktoren Profitabilität, Produktivität und Qualität auswirkt und zudem von der Performance ergänzt wird, welche Effizienz, Effektivität und geringer Kosten bedeutet [Sp14]. [GM14] erläutern den IT-Wertbeitrag als Resultat aus den im Unternehmen eingesetzten IT-Applikationen, -Prozessen und dort produzierten IT-Services sowie -Projekten, welches durch

organisatorische Aspekte, dem Mehrwert des menschlichen Wirkens, ergänzt werden [GM14].

2015 – [Af15] hat ergänzend hierzu die Erläuterung, dass der IT-Wertbeitrag als wissenschaftliches Ziel besitzt, den durch IT-Investitionen geschaffenen Wert gesamthaft zu erfassen und richtig zuzuordnen [Af15]. Weiter beschreibt [Ba15], dass der Wert der IT analog zur betriebswirtschaftlichen Definition des Wertbegriffs als Input-Output-Transformationsprozess aufgefasst werden kann, welcher in drei Komponenten beschrieben werden kann: IT-Ressourcen, IT-Leistungserstellungsprozess und dem IT-Output bzw. IT-Ziel. In der Konklusion werden demzufolge nicht nur die Ergebnisse des IT-Leistungserstellungsprozesses dem IT-Wertbeitrag zugeordnet, sondern auch die Mittel, die zu dessen Erzielung eingesetzt werden [Ba15]. [Ki15b] erläutert in [Ki15a] den Wertbeitrag der IT in der Entstehung durch niedrige Kosten in den Geschäftsprozessen, welche den Mitbewerbern gegenübergestellt werden in Form von Benchmarking. Er beruft sich zudem auf die Aussagen von [Kr10] welche sich auf den Wertbegriff beziehen, dass Wert und Nutzen einer IT-Investition unter den gegebenen Aspekten aus Sicht eines Wirtschaftssubjektes als subjektive Zusammenfassung von positiven und negativen Zielbeiträgen definiert ist, welche den Eigenschaften, Attribute einer IT-Investition oder deren Effekt auf die Diskurswelt stiften. Hierbei wird jedoch kritisch das Produktivitätsparadoxon angemerkt, welches besagt, dass kein positiver Zusammenhang zwischen IT-Investitionen und der Produktivität auf volkswirtschaftlicher Ebene existiert. [Ki15b] bezieht sich ebenfalls auf die Erläuterung von [Sp14], welche sich beide auf die Erläuterung von [HM08] beziehen. [Ki15b] ergänzt hierbei jedoch, dass die IT-Investitionen auf das Kerngeschäft des Unternehmens besonders wirken und hierdurch das (Business-) IT-Alignment versbessern [Ki15a] [Ki15b]. In der Erläuterung von [Sa15a], welche sich auf Zeiner (2009) bezieht, wird verdeutlicht, dass der IT-Wertbeitrag sich auf alle Ebenen der Wertschöpfung eines Unternehmens durchzieht sowie traditionell lediglich Aspekte der Effizienzsteigerung wie Steigerung der Produktivität und Reduktion der Kosten in den Fachabteilungen beleuchtet werden, in den Erwartungen der Geschäftsleitung jedoch oftmals Beiträge zu strategischen Ausrichtungen und verbesserter Entscheidungsfindung miteinbezogen werden [Sa15a].

2016 – [Kr16b] erläutern den Wertbeitrag der IT in dem, sämtlichen unterstützenden Auswirkungen des IT-Einsatzes auf die Ziele der Unternehmung wie z.B. die Reduktion von Kosten, Erhöhung der Produktivität, Erreichung von Wettbewerbsvorteilen und weiteren Erfolgsmessgrößen, welche sich sowohl auf die Effektivität und der Effizienz des Unternehmens beziehen, auswirken. Das IT-Alignment stellt dabei ein Instrument der IT-Governance dar, welches den effizienten IT-Einsatz innerhalb des Unternehmens sicherstellt [Kr16a] [Kr16b].

2017 – Im Beitrag von [Fr17] wird der IT-Wertbeitrag als Steigerung von finanzorientierten Parametern, intermittierend über Prozessoptimierungen oder zumindest als Wahrnehmung im Unternehmen erläutert [Fr17]. Weiter erläutert [Ki17b], dass eine effektive sowie effiziente IT-Unterstützung der Geschäftsprozesse zur Wertsteigerung des Unternehmens in erheblichem Maße beiträgt und hierbei innerhalb erfolgreicher

Unternehmen Standards eingesetzt werden in Bezug auf Hardware, Software sowie bei Prozessen, Datenstrukturen und Outsourcing (welches selektiv eingesetzt wird). Der Wert des Unternehmens wird zudem durch wertschaffende Investitionen erhöht [Ki17a] [Ki17b].

2020 - [PKS20] fassen die vielen existenten Definitionen des IT-Wertbeitrags zu einer Erläuterung in zwei Paradigmen zusammen. Im ersten konzentriert sich der Wertbeitrag der IT auf die Faktoren Kosten, Produktivität, Rentabilität und Effizienz, welche zu einem Wettbewerbsvorteil führen. Das zweite Paradigma betrachtet den IT-Wertbeitrag aus einer Ressource basierten Sichtweise in dem die IT indirekt zu einem Wettbewerbsvorteil beiträgt, in dem die Ressourcen des Unternehmens, IT-Verfahren, Vermögenswerte und IT-Prozesse genutzt werden [PKS20].

2021 - In dem Beitrag von [Se21] wird der Wertbeitrag der IT durch sechs Dimensionen und Charakteristiken erläutert, die den Einfluss der IT auf das Unternehmen verdeutlichen [Se21].

Zusammenfassend erläutert sich der Wertbeitrag der IT in finanziellen, Leistungs-, aber auch strategischen Einflüssen auf das Unternehmen, welche eine Wirkung auf dessen Erfolg besitzen. Die Erläuterungen und das eben beschriebene können zudem abgeleitet werden, wie sich der IT-Wertbeitrag äußert. Auf der einen Seite in finanzielle Faktoren wie die Kostenreduktion oder die Wirkung von Investitionen der IT auf den Erfolg des Unternehmens. Auf der anderen Seite äußert sich der IT-Wertbeitrag in Leistungseinflüssen die jedoch auch mit finanziellen Faktoren wie Investitionen zu tun haben. Hierbei äußert sich der Wertbeitrag in Produktivitätssteigerung und Prozessoptimierungen. Strategische Einflüsse bei Entscheidungen und Zielsetzungen der IT äußern sich zudem in einer schnelleren Erreichung der Ziele durch die IT oder auch dem Herausarbeiten von Wettbewerbsvorteilen. Der größte Faktor, der mit der Äußerung des IT-Wertbeitrags zu tun hat, bezieht sich auf die Effizienz und die Effektivität in allen genannten Einflüssen (finanziell, leistungsorientiert und strategisch), die sich im Wertbeitrag der IT zum Unternehmenserfolg äußern.

2.2 Ansätze der Messbarkeit

Vorerst können die Ansätze der Messbarkeit in vier Kategorien nach [PKS20] eingeordnet werden – *Finanzielle Methoden zur Messung* (z.B. Economic Value Added), *Operationale Perfomance Messmethoden* (Produktivität), *Prozessmessmethoden* (Prozesseffizienz/-effektivität) und *Methoden der Wahrnehmung zur Messung* (Verfügbarkeit von Informationen) [PKS20]. Die Ansätze werden versucht in die vier Kategorien zu klassifizieren. Da die Kategorien sich überschneiden können wird lediglich der Ansatz mit seiner Ansatzcharakteristik erläutert.

Einen ersten sehr ausführlichen Ansatz liefert Bearingpoint im Beitrag von [Sp14], der *Prozessmessmethoden*, *operationale Performance Messmethoden* sowie *Methoden der Wahrnehmung zur Messung* beinhaltet. Hierbei wurden über bestimmte Fragen, fünf

Wertschöpfungssektoren abgeleitet, welche mit einem prozentualen Faktor gewichtet wurden. Ergänzt werden diese durch IT-Maßnahmen zur Erreichung strategischer Ziele, welche in Handlungsfelder mit ebenfalls prozentualen Gewichtungsfaktoren der Wirkung auf die fünf Wertschöpfungssektoren festgelegt wurden. Die Wertschöpfungssektoren und Handlungsfelder werden in einer Matrix mit ihren Gewichtungsfaktoren aufgespannt. Das Unternehmen berechnet anschließend welchen prozentuellen Grad des Wertschöpfungssektors diese erreichen. Der vom Unternehmen berechnetet Faktor (max. 100 %) wird dann mit dem Faktor des Wertschöpfungssektors multipliziert. Die Ergebnisse aus den fünf Sektoren werden dann addiert und ergeben den Reifegrad des momentan erreichten IT-Wertbeitrags des Unternehmens [Sp14].

Der Ansatz von [GM14] vorgestellt und von Frauenhofer ISST entwickelt, befindet sich kategorisch in der *operationalen Performance Messung*. Das Modell zur Wertbeitragsmessung nennt sich ITEM – IT-Evaluation Management. Dieser basiert auf einer Balanced Scorecard und stellt einen Ansatz zur Steuerung der Unternehmens-IT dar, welcher sich an deren Wert orientiert. Hierbei werden Kenn- und Messzahlensysteme vorgestellt, die den Nutzen der IT quantifizieren. Die ITEM-Methodik geht hierbei in fünf Schritten vor. Der Wertbeitrag der IT wird dabei im Durchlaufen der Prozesse ermittelt [GM14].

[Af15] liefert einen Ansatz zur *Wahrnehmungsmethode* des IT-Wertbeitrags. Dieser basiert auf der *Prospect-Theory*, einem Ansatz aus der Verhaltensökonomie, welche zur deskriptiven Analyse der menschlichen Wahrnehmung unter Risiko verwendet wird. Werden die Ergebnisse aus dieser Theorie auf IT-Entscheidungen angewandt, so wird der wahrgenommene IT-Wertbeitrag zum Unternehmenserfolg quantifiziert und analysiert [Af15].

Im Beitrag von [Ba15] wird ein Referenzmodell zum Wertbeitrag der IT entwickelt, in welchem Rahmen auch auf die Feststellung dessen eingegangen wird. Hierbei wird der Wertbeitrag in strategische, transaktionsbezogene, informationsbezogene, wertsichernde und transformationsbezogene Wertbeiträge kategorisiert, über welche zudem die einzelnen Wertbeiträge festgestellt werden können. Es werden weiter verschieden Wirkungsmodelle für die Generierung des IT-Wertbeitrags auf Outcome-Ebene vorgestellt. Zudem wird der IT-Wertbeitrag quantifiziert [Ba15]. Kategorisch ziehen sich diese über die *finanziellen, operationalen Performance- und Prozessmessmethoden*.

In dem Beitrag von [Sa15b], welcher in die Kategorie von *operationalen Performancemessmethoden* eingeordnet werden kann, wird ein weiterer Ansatz präsentiert, welcher über ein sog. ValueBoard eine strukturierte Methode für Messungen und Messgrößen zur Darstellung des „ex-post“ Wertbeitrags von IT-Projekten vorschlägt. Das ValueBoard orientiert sich an Best Practise Methoden, welches auf den Ebenen des Business Engineering aufsetzt und die Ergebnisse entsprechend der Prinzipien industrieller Regelkarten visualisiert. Hierdurch wird eine ergebnisorientierte Sicht auf den tatsächlichen Nutzen von IT-Investitionen realisiert [Sa15b].

[PJ16] geben einen sehr allgemeinen Ansatz zur Messung des IT-Wertbeitrags, indem der Wertbeitrag anhand von klassischen Methoden der BWL ermittelt wird. Diese belaufen

sich auf die Kostenebene und Schlagen zudem Methoden des Benchmarkings vor, um den Wertbeitrag der IT gegenüber anderen Unternehmen festzustellen. Generell wird hierbei jedoch keine explizite Methode vorgeschlagen, den IT-Wertbeitrag zu messen, es wird mehr auf den generellen Ansatz den Wertbeitrag überhaupt zu messen eingegangen. Kategorisch kann hierbei von *finanziellen* und *operationalen Performancemessmethoden* gesprochen werden [PJ16].

[SSW19] beschreibt als Ansatz zur Feststellung des Wertbeitrags, dass vorerst die Wirkung des Wertbeitrags größtenteils identifiziert werden muss. Diese muss anschließend einer monetären Bewertung unterzogen werden, in dem die identifizierten Potenziale mit möglichen Umsatzeffekten bewertet werden. Hierzu werden Methoden der statischen sowie dynamischen Investitionsrechnung herangezogen, um die IT-Investitionen zu bewerten. Für die dynamischen Rechnungen werden die Kapitalwert- oder Annuitätenmethode sowie der Finanzplan des Unternehmens genutzt. Die Vorgehensweise von [SSW19] kann deshalb den *finanziellen Messmethoden* zugeordnet werden [SSW19].

In dem Beitrag von [TDM20] werden finanzielle und nichtfinanzielle Methoden genutzt, um den IT-Wertbeitrag zu messen. [TDM20] schlagen bei finanziellen Methoden bspw. Kennzahlen bzw. Kennzahlensysteme vor, mit dem der IT-Wertbeitrag ausgerechnet werden kann. In den nichtfinanziellen Methoden wird von [TDM20] vorgeschlagen, bspw. das sog. Scoring anzuwenden [TDM20]. Dieser Beitrag bzw. Ansatz kann den *finanziellen* sowie *operationalen Performancemessmethoden* zugeordnet werden.

Die letzten beiden Beiträge von [SSW21] und [Se21] zur Wertbeitragsmessung basieren aufeinander und schlagen zur Messung des IT-Wertbeitrags einen Wertekatalog vor, welcher auf das jeweilige Unternehmen angepasst wird. Zudem wird der IT-Wertbeitrag in seine Hauptbestandteile zerlegt. Der Katalog wird anhand von vier Schritten erarbeitet. Das Modell befindet sich derzeit noch in der Entwicklung und kann zudem noch nicht in eines der vier Kategorien zugeordnet werden.

Abschließend kann gesagt werden, dass viele Ansätze zur IT-Wertbeitragsmessung existieren, jedoch keiner dieser Ansätze einen standardisierten und absoluten Lösungsansatz der Problematik liefert. Viele Ansätze basieren auf ähnlichen Vorgehensweisen, die sich dennoch voneinander unterscheiden. Besonders häufig werden Ansätze in Bezug auf finanzielle Methoden zur Messung, operationale Perfomance- und Prozessmessmethoden oder eine Mischung aus diesen angewandt. Eine klare und definierte Aussage welche Methoden bestmöglich für die vielseitigen Bereiche der IT geeignet sind, geht aus den Ansätzen nicht hervor.

3 Fazit

Diese Arbeit schafft ein Verständnis des Wertbeitrags der IT zum Unternehmenserfolg und stellt dar, wie sich dieser äußert. Zudem werden Ansätze aufgezeigt, wie der Wertbeitrag der IT zum Unternehmenserfolg messbar gemacht werden kann. Dabei

wurden die Forschungsfragen Q1 und Q2 beantwortet. Der Wertbeitrag der IT bildet im Kern den Mehrwert, den die IT zum Unternehmenserfolg leistet. Dabei sind IT-Ressourcen Treiber, welche die Effizienz sowie die Wettbewerbsfähigkeit eines Unternehmens beeinflussen. Hierbei spielen Produktivitätssteigerung und Kostensenkung durch die IT ebenfalls eine Entscheidende Rolle für den Wertbeitrag der IT. Der IT-Wertbeitrag äußert sich in finanziellen, Leistungs- und strategischen Einflüssen, die die IT auf das Unternehmen hat, welche eine bestimmte (positive oder negative) Wirkung zum Erfolg ausüben. Zudem existieren mehrere Ansätze zur IT-Wertbeitragsmessung, welche in die vier Kategorien *finanzielle Messmethoden*, *operationale Perfomancemessmethoden*, *Prozessmessmethoden* und *Messmethoden der Wahrnehmung* eingeteilt werden können. Besonders häufig treten die ersten drei Kategorien auch in Verbindung miteinander auf. Kritisch muss bei dieser Arbeit jedoch betrachtet werden, dass diese zwar ein gutes Verständnis des IT-Wertbeitrags und Ansätze dessen Messung liefert, jedoch hierbei nicht die Tiefe abgedeckt, um eine konkrete Messmethode ableiten zu können und einen standardisierten Lösungsvorschlag der genannten Problematik zu liefern. Auch werden die einzelnen Teilbereiche der IT ((*Business) -IT-Alignment*, *IT-Governance*, *IT-Management* und *IT-Controlling*) zu wenig beleuchtet, um Rückschlüsse auf die Relevanz dieser Bereiche im Kontext des Wertbeitrags zu schließen. Überleitend hierzu ergibt sich der konkrete Forschungsbedarf, die Relevanz des Wertbeitrags der IT im Kontext der einzelnen IT-Bereiche zu untersuchen, um Rückschlüsse auf die Wichtigkeit derer in Bezug auf den IT-Wertbeitrag ermöglichen zu können. Weiter kann die Relevanz des IT-Wertbeitrags in Bezug auf die Wirtschaftsinformatik untersuchten werden. Ein weiterer Forschungsvorschlag ist, die Ansätze inhaltlich zu analysieren und hierdurch zu identifizieren wie Unternehmen gezielt den IT-Wertbeitrag messen können und wie die einzelnen Ansätze zusammengesetzt werden müssen, um eine bestmögliche Messung im Generellen zu ermöglichen.

Literaturverzeichnis

[Af15] Afflerbach, P.: The Business Value of IT in Light of Prospect Theory. Business & Information Systems Engineering 5/57, S. 299–310, 2015.

[BDK21] Bächle, M. A.; Daurer, S.; Kolb, A.: Einführung in die Wirtschaftsinformatik. De Gruyter, 2021.

[Ba15] Bartsch, S.: Der Gegenstand des Wertbeitrags der IT. In (Bartsch, S. Hrsg.): Ein Referenzmodell zum Wertbeitrag der IT. Springer Fachmedien Wiesbaden, Wiesbaden, S. 139–208, 2015.

[Br09] Brocke, Jan vom. et al.; "RECONSTRUCTING THE GIANT: ON THE IMPORTANCE OF RIGOUR IN DOCUMENTING THE LITERATURE SEARCH PROCESS" (2009). ECIS 2009 Proceedings. 161.

[Fr17] Franke, M. R.: Dynamic Capabilities und Wettbewerbsfähigkeit durch Cloud Computing. Springer Fachmedien Wiesbaden, Wiesbaden, 2017.

[GM14] Gadatsch, A.; Mayer, E.: Masterkurs IT-Controlling. Springer Fachmedien Wiesbaden, Wiesbaden, 2014.

[Ki17a] Kischporski, M. Hrsg.: EDI – Digitalisierung und IT-Wertbeitrag konkret umgesetzt. Springer Fachmedien Wiesbaden, Wiesbaden, 2017.

[Ki15a] Kischporski, M. Hrsg.: Elektronischer Rechnungsdatenaustausch mit E-Invoicing. Springer Fachmedien Wiesbaden, Wiesbaden, 2015.

[Ki17b] Kischporski, M.: (IT-)Wertbeitrag und Business-/IT-Alignment. In (Kischporski, M. Hrsg.): EDI – Digitalisierung und IT-Wertbeitrag konkret umgesetzt. Springer Fachmedien Wiesbaden, Wiesbaden, S. 43–44, 2017.

[Ki15b] Kischporski, M.: Wertbeitrag – Optimierungsvarianten. In (Kischporski, M. Hrsg.): Elektronischer Rechnungsdatenaustausch mit E-Invoicing. Springer Fachmedien Wiesbaden, Wiesbaden, S. 43–54, 2015.

[Kr16a] Kröschel, I. Hrsg.: Architekturbasiertes IT-Alignment. Springer Fachmedien Wiesbaden, Wiesbaden, 2016.

[Kr16b] Kröschel, I.: Terminologische und konzeptionelle Grundlegung. In (Kröschel, I. Hrsg.): Architekturbasiertes IT-Alignment. Springer Fachmedien Wiesbaden, Wiesbaden, S. 25–114, 2016.

[Ly20] Lynn, T. et al. Hrsg.: Measuring the Business Value of Cloud Computing. Springer International Publishing, Cham, 2020.

[PKS20] Pathak, S.; Krishnaswamy, V.; Sharma, M.: Impact of IT practices and business value of IT measurement. International Journal of Productivity and Performance Management 4/69, S. 774–793, 2020.

[PJ16] Pfitzinger, B.; Jestädt, T.: IT-Betrieb. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.

[Sa15a] Samulat, P.: Messkonzept für ein benutzerzentrisches Kennzahlensystem zur Darstellung des Wertbeitrags der IKT. 2194-105X, 2015.

[Sa15b] Samulat, P., (2015). Methode für Messungen und Messgrößen zur Darstellung des „Ex-post“ Wertbeitrages von IT-Projekten. HMD Praxis der Wirtschaftsinformatik: Vol. 52, No. 4. Springer. (S. 624-635).

[SG20] Schnabl, A.; Grechenig, T.: Benefit Evaluation in IT Project Portfolio Management at Austrian Financial Service Providers: Proceedings of the 2020 the 3rd International Conference on Computers in Management and Business. ACM, New York, NY, USA, S. 290–295, 01312020.

[SSW19] Schütte, R.; Seufert, S.; Wulfert, T.: Das Wertbeitragscontrolling als Anreicherung bestehender Vorgehensmodelle des Software Engineering. 1617-5468, 2019.

[SWW21] Seufert, Sarah; Wulfert, Tobias; and Wernsdörfer, Jan Eric, "TOWARDS A REFERENCE VALUE CATALOGUE FOR A COMPANY-SPECIFIC ASSESSMENT OF THE IT BUSINESS VALUE – PROPOSING A TAXONOMY TO SELECT IT IMPACTS FROM EXISTING CATALOGUES" (2021). ECIS 2021 Research-in-Progress Papers. 34.

[Se21] Seufert, S. et al.: A Literature-based Derivation of a Meta-framework for IT Business Value: Proceedings of the 23rd International Conference on Enterprise Information Systems. SCITEPRESS - Science and Technology Publications, 2021.

[Sp14] Spörrer, S.: Business Continuity Management. Springer Fachmedien Wiesbaden, Wiesbaden, 2014.

[TMD20] Tallon, P. P.; Mooney, J. G.; Duddek, M.: Measuring the Business Value of IT. In (Lynn, T. et al. Hrsg.): Measuring the Business Value of Cloud Computing. Springer International Publishing, Cham, S. 1–17, 2020.

[TND14] Termer, F.; Nissen, V.; Dorn, F.: Grundlegende Überlegungen zur Konzeption einer multiperspektivischen Methode der Messung des IT-Wertbeitrags. In: Tagungsband Multikonferenz Wirtschaftsinformatik, S. 2271-2283, 2014.

[Wi14] Wiedenhofer, A.: Flexibilitätspotenziale heben – IT-Wertbeitrag steigern. Springer Fachmedien Wiesbaden, Wiesbaden, 2014.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 2/26, S. xiii–xxiii, 2002.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1641

Überblick über die Ansätze zur Geschäftsprozess-automatisierung

Eine systematische Literaturanalyse zum State-of-the-Art

Remzi Kültür[1], Helmut Beckmann[2]

Abstract: Aufgrund der zunehmenden Bedeutung der Automatisierung von Geschäftsprozessen, werden Unternehmen gezwungen alles zu automatisieren, was unter Berücksichtigung wirtschaftlicher Faktoren rentabel ist. Allerdings fällt es Unternehmen schwer, bei so vielen Ansätzen zur Automatisierung von Geschäftsprozessen einen ganzheitlichen Überblick zu bewahren. Darüber hinaus ist zu beachten, dass den Unternehmen das gewisse Knowhow über die Existenz dieser Ansätze fehlt. Daher ist das Ziel dieser Arbeit, die Ansätze zur Automatisierung von Geschäftsprozessen anhand der Literatur zu ermitteln. Dadurch soll ein Überblick über diese Ansätze verschafft werden, um den Mangel an Transparenz und Wissen im Bereich der Automatisierung von Prozessen im Unternehmen zu bewältigen.

Keywords: Geschäftsprozessautomatisierung, Ansatz, Automatisierung von Geschäftsprozessen, Technologe

1 Einleitung

Die Bedeutung der Geschäftsprozessautomatisierung hat in den letzten Jahren deutlich zugenommen und ist eines der am häufigsten wiederkehrenden Themen in der Geschäftswelt. Die Wettbewerbsfähigkeit zwingt die Unternehmen dazu, ihre Produkte oder Dienstleistungen effizient und effektiv zu liefern. Sie müssen nicht nur einen angemessenen Wert bieten, sondern dies auch schneller und mit höherer Qualität tun. Dieser agile Kontext führt dazu, alles zu automatisieren, was sich unter Berücksichtigung wirtschaftlicher Faktoren lohnt zu automatisieren, um den Schwerpunkt auf den Wert zu legen und gleichzeitig die Bearbeitungszeiten, Fehlerraten und die Prozessleistung im Allgemeinen zu optimieren. Da die Automatisierung von Geschäftsprozessen seit vielen Jahren thematisiert wird, existieren bereits verschiedene Ansätze, die in den einzelnen Bereichen des Unternehmens eingesetzt werden können. Diese Ansätze können durch konkrete Technologien umgesetzt werden, um wichtige Merkmale, wie z.B. die Effizienz der Automatisierung zu steigern und generell die Automatisierung zu beschleunigen. Allerdings fällt es Unternehmen schwer bei so vielen Möglichkeiten zur Automatisierung

[1] Hochschule Heilbronn, Fakultät Wirtschaft und Verkehr, Anschrift, Max-Plank-Str. 39, 74081 Heilbronn, rkueltuer@stud.hs-heilbronn.de
[2] Hochschule Heilbronn, Fakultät Wirtschaft und Verkehr, Anschrift, Max-Plank-Str. 39, 74081 Heilbronn, helmut-beckmann@hs-heilbronn.de

doi:10.18420/inf2022_141

ihrer Geschäftsprozesse einen holistischen/ganzheitlichen Überblick zu behalten. Andererseits sind sich Unternehmen nicht bewusst, oder ihnen fehlen das bestimmte Knowhow über die Existenz dieser Ansätze. Diese Unklarheit und Unwissenheit können dazu führen, dass sich Unternehmen vom Wettbewerb zurückziehen und somit den Erfolg des Unternehmens beeinflussen.

Aus der oben beschriebenen Problemstellung stellt sich für diese Arbeit folgende Forschungsfrage: *„Welche Ansätze stehen Unternehmen zur Verfügung, die eine Automatisierung ihrer Geschäftsprozesse anstreben?“* Die Analyse der Forschungsfrage soll Kenntnis darüber geben, welcher aktuelle Wissenstand hinsichtlich der Geschäftsprozessautomatisierung vorhanden ist. Entsprechend werden Ansätze zur Automatisierung von Prozessen ermittelt. Damit soll ein Überblick über diese Ansätze verschafft werden, um den Mangel an Transparenz und Wissen im Bereich der Automatisierung von Prozessen im Unternehmen zu bewältigen.

Zu diesem Zweck werden in Kapitel 2 zunächst erläutert, welche Automatisierungsstufen ein Geschäftsprozess erreichen kann. Anschließend erfolgt in Kapitel 3 die angewandte Methodik zur Literaturanalyse. Infolgedessen werden in Kapitel 4 die Ergebnisse dieser Arbeit vorgestellt. Abschließend folgt in Kapitel 5 die Zusammenfassung der Ergebnisse sowie ein Ausblick über zukünftige Forschungsarbeiten.

2 Grundlagen und Begriffe

Stufen der Automatisierung

Die Ausprägung / der Reifegrad der Automatisierungsansätze kann in fünf Stufen unterteilt werden [Gae20] (vgl. Tabelle 4). Diese Stufen unterscheiden sich zwischen den Teilen der Aufgabenerfüllung bzw. Entscheidungsfindung, die von Menschen durchgeführt werden und denen, die von Systemen automatisiert werden (*wer automatisiert* und *was wird automatisiert*).

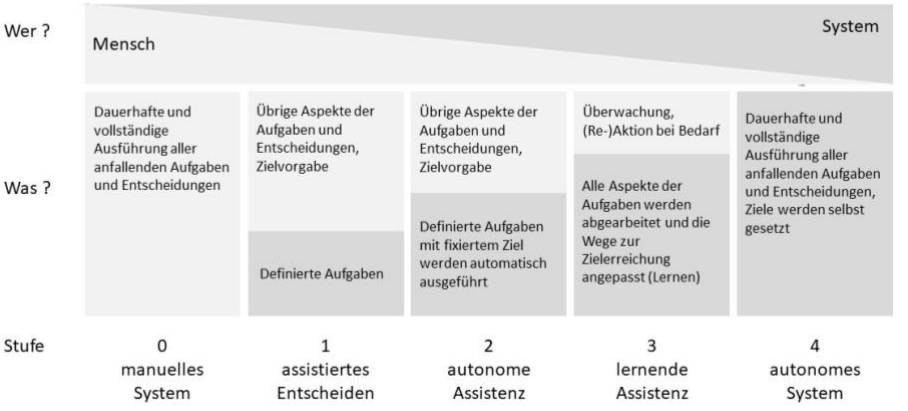


Abb. 1: Stufen der Automatisierung (eigene Darstellung, Anlehnung an Gärtner [Gae20]

3 Methodik zur Literaturanalyse

Verfahren

Für die Ermittlung der Ergebnisse wurde eine strukturierte Literaturanalyse nach Fettke [FE06] durchgeführt. Damit bei der Suche nach relevanter Literatur eine Vollständigkeit erbracht werden kann, wurde zusätzlich eine Vor- und Rückwärtssuche nach Webster und Watson [WW02] verwendet. Des Weiteren werden die Ergebnisse anhand einer Konzeptmatrix dargestellt.

4 Ergebnisse

Identifizierte Ansätze zur Geschäftsprozessautomatisierung

Auf Basis der Literaturanalyse wurden insgesamt acht Ansätze identifiziert, die ihre Relevanz im Bereich der Geschäftsprozessautomatisierung belegen. Diese Ansätze sind: Workflow, Robotic Process Automation (RPA), Intelligent Process Automation (IPA), Hyperautomation, Makro, Scripting, Screen-Scraping und Smart Contracts.

Literatur	Identifizierte Ansätze							
	Workflow	RPA	IPA	Hyper-automation	Makro	Screen Scraping	Scripting	Smart Contracts
[Sc18]								
[Ch17]	x	x	x					
[Me17]	x							
[Sh19]		x	x					
[Du21]	x							
[Gae20]		x	x					
[Fe06]								
[WW02]								
[Ga20]	x							
[Ra20]	x	x						
[Cz18]		x			x	x	x	
[Fe21]								
[Du18]						x		
[Me18]		x						

[Ri21]		x						
[Ji17]		x		x				
[Ta20]		x	x		x		x	
[La21]		x	x					
[Ma20]	x	x	x			x		
[Sm19]		x	x		x	x	x	
[Ha21]		x		x				
[He21]		x	x	x				
[Ha20]		x			x	x	x	
[Kh21]								x
[Ma19]								x
[Eg20]	x	x						x
[Bh21]								x

Tab. 1: Konzeptmatrix: identifizierte Ansätze (eigene Darstellung)

Die Konzeptmatrix (vgl. Tabelle 4) gibt einen Überblick über die ermittelten Ansätze anhand der Literaturanalyse. Insgesamt wurden acht Ansätze identifiziert, die für die Automatisierung von Geschäftsprozessen in Anspruch genommen werden können. Anhand Tabelle 4 wurde eine Übersicht über die Verteilung der unterschiedlichen Ansätze generiert (vgl. Abbildung 2). RPA tritt in der untersuchten Literatur am häufigsten auf (16-mal). IPA hingegen wird als Erweiterung von RPA angesehen und wird nur zusammen mit RPA erwähnt (8-mal) [Ch17]. Trotz ihres bereits hohen Reifegrades (Workflow Management Coalition (WFMC) existiert seit 1993) [Du21] werden weiterhin Artikel über Workflows veröffentlicht (7-mal). Dies ist ein Indikator dafür, dass Workflows noch nicht ihr volles Potenzial erreicht haben. Der Ansatz Hyperautomation kommt nur 3-mal in der Literatur vor. Das kann dadurch begründet werden, dass dieser Ansatz ein relativ neuer Ansatz ist. Jedoch wurde dieser Ansatz in der Studie „Gartner Top 10 Strategic Technology Trends for 2020“ als der Top-Trend der strategischen Technologien des Jahres gewählt [He21]. Das deutet daraufhin, dass dieser Ansatz bei Unternehmen zunehmend an Popularität gewinnt und in den kommenden Jahren eine wichtige Rolle im Bereich der Automatisierung einnehmen wird. Makros, Screen-Scraping und Scripting hingegen, werden deutlich weniger in der aktuellen Literatur erwähnt. Allerdings dienen diese Ansätze in der State-of-the-Art primär als Basis für andere Ansätze [Cz18]. Smart Contracts tauchen in der Automatisierungsliteratur ebenso selten auf wie Makros und Scripting, hierbei lässt sich die geringe Anzahl an Veröffentlichungen durch den noch unerforschten Anwendungsbereich begründen [Bh21].

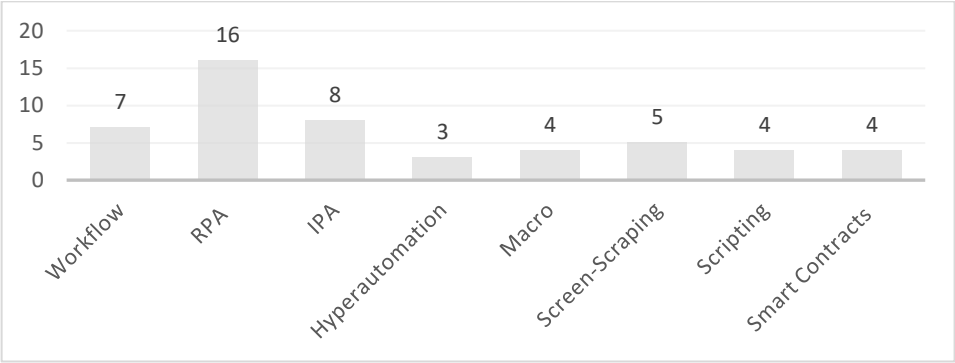


Abb. 2 Verteilung der Ansätze in der identifizierten Literatur (eigene Darstellung)

4.1 Workflow (Automatisierungsstufe 1/2)

Unter dem Begriff Workflow wird die vollständige oder teilweise Automatisierung eines Geschäftsprozesses verstanden. Der Einsatz von Workflows umfasst Zeit-, Funktions- und Ressourcenspezifikationen. Sie dient zur automatischen Steuerung eines Arbeitsablaufes, welches sich auf der operativen Ebene befindet. Workflows beinhalten eine Reihe von Regeln und Kriterien, welche für die Ausführung erforderlich sind. Ein Workflow beschreibt also, welche Aufgaben von wem und wie ausgeführt werden. [Ga20] Für die Ausführung von Workflows werden in der Literatur zwischen prozessorientierten domänenspezifischen Informationssystemen und Geschäftsprozessmanagementsystemen (engl. Business Process Management Systeme (BPMS)) unterschieden [Sc18] [Ra20]. Zu den domänenspezifischen Informationssystemen, welche auch als Anwendungssysteme bezeichnet werden [Cz18], gehören unter anderem Enterprise Resource Planning (ERP)-, Customer-Relationship-Management (CRM)- und Supply Chain Management (SCM)-Systeme. Diese Systeme sind unternehmensübergreifend integriert und können durch ein EAI-System (Enterprise Application Integration) integriert werden, um einen Austausch von Informationen zwischen Systemen zu ermöglichen [Fe21]. Im Gegensatz dazu werden BPMS in vier verschiedene Sub-Systeme eingeordnet, in denen der Grad der Unterstützung erfasst und geprüft wird, wie sich die Systeme hinsichtlich der Ausrichtung ihrer Prozesse oder Daten voneinander unterscheiden [Du21]. Das erste Subsystem nennt sich *Groupware-System* und bietet den Benutzern die Möglichkeit, für den Austausch von Dokumenten und Informationen sowie eine direkte Kommunikation. Das *Ad-hoc-Workflowsystem* ermöglicht die Echtzeiterstellung und -änderung eines definierten Prozesses. Die strikte Einhaltung von explizit definierten Prozessbeschreibungen wird als *Produktions-Workflow-System* bezeichnet. Für eine Automatisierung von dezentralen Geschäftsprozessen ist diese Möglichkeit ebenfalls ein wichtiger Faktor. Das letzte Sub-System wird *Case-Management-System* bezeichnet und dient zur Unterstützung der nicht detaillierten und unvollständig spezifizierten Prozesse. Bei den oben genannten Erscheinungsformen von Workflows sind weiterhin manuelle Tätigkeiten wie die

Vorbereitung von Eingabedaten und deren Eingabe in Formularfelder erforderlich, was sich negativ auf die Effizienz und Qualität auswirkt [Gae20] [Cz18]. Durch die Integration weiterer Technologien kann jedoch eine höhere Automatisierungsstufe erreicht werden. Beispielweise wurde das System in den jüngsten Jahren durch die Radio Frequency Identification (RFID) Technologie erweitert, welche die detaillierte Verfolgung von Objekten im Herstellungsprozess ermöglicht. Dementsprechend befinden sich Workflows in den Stufen 1 und 2.

4.2 Robotic Process Automation (RPA) (Automatisierungsstufe (2/3)

Für die Automatisierung der bestehenden, manuellen Tätigkeiten wird RPA eingesetzt. Gegenwärtig ist der Einsatz von Robotic Process Automation (RPA) für Unternehmen von großer Bedeutung. RPA ist die Reaktion der Industrie auf die enorme Masse an manuellen Aufgaben bzw. Aktivitäten, die täglich, wöchentlich, oder monatlich von Menschen ausgeübt werden [Me18]. Daher konzentrieren sich die RPA-Lösungen nur auf einzelne Teilprozesse oder Aktivitäten. Laut unterschiedlicher Literatur befolgt RPA das Ziel, die wiederholenden arbeitsintensiven Aufgaben zu beseitigen [Ch17] [Ri21]. Auch wenn der erste Gedanke, der bei dieser Automatisierungstechnik aufkommt, sich auf Robotern bezieht, handelt sich hierbei um eine installierbare Software [Sh19]. Demnach ist RPA ein Software-Paradigma, welches den Software-Maschinen ermöglicht, mit Informationssystemen (wie z.B. ERP-Systeme) über deren Benutzerschnittstellen zu interagieren [Ji17]. Grundsätzlich lässt sich RPA in der Literatur in zwei Arten unterscheiden, in *Attended* (Beaufsichtigte) RPA und in *Unattended* (unbeaufsichtigte) RPA [Ta20]. Beim Attended RPA interagiert die Software mit einem Anwender. Beispielsweise kann Attended RPA im Rechnungswesen & Controlling für das Monats-Reporting eingesetzt werden, indem der Roboter durch den Anwender dazu veranlasst wird, die aktuellen Monatszahlen aus den Systemen zu extrahieren und nach dem Abschluss zu verdichten [Ta20] [La21]. Unattended RPA hingegen automatisiert Aktivitäten ohne menschliche Beteiligung, d. h. das System reagiert auf bestimmte Ereignisse, z. B. ein Kunde schickt eine Kündigungsmitteilung per Mail [La21]. Wird das Monats-Reporting nun nach einem vordefinierten Zeitplan veranlasst – und fällt so die Interaktion mit dem Anwender weg – handelt es sich ebenfalls um Unattended RPA. Der Einsatz von RPA im Unternehmen bringt viele Vorteile mit sich. Dieser Ansatz kann zu Kosteneinsparungen in den einzelnen Prozessen führen [Ma20]. Darüber hinaus kann die Qualität der Prozesse durch eine geringere Fehlerquote einer Software im Vergleich zu einem Menschen verbessert werden.

4.3 Intelligent Process Automation (IPA) (Automatisierungsstufe 3/4)

IPA ist die Weiterentwicklung von RPA, bei der künstlichen Intelligenz eingesetzt wird, um das vom Menschen abhängige Training zu minimieren [Ch17]. Diese Kombination kann die traditionelle Entscheidungsfindung übertreffen, um die betriebliche Effizienz, Entscheidungsqualität und Systemzuverlässigkeit zu erhöhen. RPA

wird primär für Prozesse eingesetzt, welche einen wiederkehrenden, regelbasierten Charakter und strukturierte Daten (z.B. Tabellen, Datenbanken) aufweisen. Die KI hingegen verfügt über kognitive Fähigkeiten, um das menschliche Verhalten zu emulieren und unstrukturierte Daten (z.B. Video, Bilder) durch Einsatz von KI-Technologien wie Optical Character Recognition (OCR) (Technologie zur Zeichenerkennung) oder Natural Language Processing (NLP) (Verarbeitung von natürlichen Sprachen) zu verarbeiten [La21]. Durch das Zusammenspiel von RPA und KI ist die Software in der Lage, menschliche Eingriffe zu minimieren und aufgrund der gewonnenen Erkenntnisse eigenständig Entscheidungen zu treffen [Ta20]. Ein Roboter könnte zum Beispiel eine eingehende E-Mail eines Kunden mit OCR lesen und dann die NLP-Technologie verwenden, um die Nachricht auszuwerten und zu überprüfen. Hierbei kann der Roboter entscheiden, ob es sich um eine neutrale Anfrage zu einem Produkt oder um eine Reklamation handelt und kann die Anfrage entsprechend klassifizieren sowie an den zuständigen Service-Mitarbeiter weiterleiten [La21]. Außerdem können sogenannte Chatbots (Dialogsysteme) eingesetzt werden, die auf gesprochene oder geschriebene Nachrichten anhand generativer Modelle (Machine Learning Ansatz) wie Long-Short-Term Memory (LSTM), antworten können [Gae20]. Demnach können IPAs je nach verwendeter Technologie sich in der Stufe 3 und 4 befinden.

4.4 Hyperautomation (Automatisierungsstufe 1/2/3/4)

Hyperautomation ist, einfach ausgedrückt, eine Verschmelzung von mehreren Ansätzen und Technologien der Automatisierung (z.B. BPMS, RPA und IPA) und künstlicher Intelligenz (z.B. OCR). Diese Kombination verbessert die menschlichen Fähigkeiten und versetzt sie in die Lage, Tätigkeiten schneller, effizienter und mit weniger Fehlern auszuführen [Ha21]. Zudem wird Hyperautomation eingesetzt, um die Digitalisierung im Unternehmen voranzutreiben (z.B. neue Technologien, neue Geschäftsmodelle) und erfordert dementsprechend Unterstützung durch die Entscheidungsträger im Unternehmen sowie die Einhaltung einer festen Roadmap, um den Ansatz langfristig und erfolgreich im Unternehmen zu etablieren. Eine weitere Zielvorgabe der Hyperautomatisierung ist es, Mitarbeiter von zunehmenden Tätigkeiten zu entlasten, um ihnen die Möglichkeit zu bieten sich auf wichtigere Aufgaben zu konzentrieren. Im Hinblick auf die Natur dieses Ansatzes, lohnt es sich nur für hoch-digitale Unternehmen Hyperautomatisierung einzusetzen. Darüber hinaus sind die Komplexität und die Realisierungskosten der Hyperautomatisierung höher als bei den restlichen Ansätzen. [He21]

4.5 Makros, Scripting, Screen Scraping (Automatisierungsstufe 1)

Neben RPA sind Makros, Scripting und Screen Scraping weitere Ansätze zur Automatisierung von Prozessen, welche kein dediziertes Backendsystem benötigen. Czarnecki und Auth [Cz18] weisen jedoch darauf hin, dass das Grundprinzip von RPA auf den technisch bestehenden Ansätzen wie Makro, Scripting und Screen Scraping basiert. RPA hat diese in der Verwendung weitestgehend abgelöst. Bei einem Makro

handelt es sich um eine kurze Sequenz vom Code, die wiederholt ausgeführt werden muss, um eine bestimmte Aufgabe zu erfüllen [Ha20]. Zu den typischen Makros gehören Tastatur- und Mauskürzeln [Ha20]. Das Scripting ist ein ähnlicher Ansatz wie der Makro-Ansatz, bei dem jedoch ein bestimmter Vorgang oder eine bzw. mehrere Aufgaben ausgeführt werden [Ha20]. Des Weiteren können Skripte in anderen Programmen und Subroutinen verwendet werden. Laut Hammer et. al. [Ha20] unterscheiden sich Makros und Scripting gegenüber RPA darin, dass RPA komplexere und anwendungsübergreifende Aufgaben ausführen können. Zudem deuten Hammer et. al. [Ha20] an, dass der Unterschied darin liegt, dass RPA gegenüber diesen beiden Ansätzen den Fokus auf die Governance legt. Der Screen Scraping Ansatz ist charakterisiert durch das Kopieren bzw. Extrahieren von Daten bzw. Informationen, welche sich auf der Benutzeroberfläche eines Computers befinden [Du18]. Der Unterschied zwischen RPA und Screen Scraping liegt darin, dass RPA mit technischen Selektoren und der Screen Scraping Ansatz mit Pixeln arbeiten, um beispielsweise einen zu drückenden Button zu finden [La21]. Somit sind RPA weitaus fortschrittlicher und bieten einen flexibleren, skalierbaren und zuverlässigeren Einsatz im Unternehmen [Ha20]. Sowohl Makros, Scripting wie auch Screen Scraping zeichnen sich durch geringe Kosten und kürzere Implementierungszeiten aus, wie auch niedrigere Komplexität. Daher eignen sich diese Ansätze für einen Einstieg in die Automatisierung. Da bei diesen Ansätzen nur definierte Aufgaben automatisiert werden, handelt es sich um die Automatisierungsstufe 1.

4.6 Smart Contracts (Automatisierungsstufe 2)

Intelligente Verträge sind digitale oder elektronische Verträge, die auf der Blockchain-Technologie basieren. Dieser Ansatz unterscheidet sich von den bisher erwähnten Ansätzen. Sie ermöglichen die rechtssichere Automatisierung von Geschäftsprozessen zwischen zwei oder mehreren Parteien [Me18]. Durch den Einsatz der Blockchain-Technologie sind diese Prozesse fälschungssicher, transparent und jederzeit verfügbar [Kh21]. Sie werden in der Regel eingesetzt, um die Ausführung einer Vereinbarung zu automatisieren, sodass alle Parteien sofort Gewissheit über das Ergebnis haben, ohne die Beteiligung eines Vermittlers oder Zeitverluste [Ma19]. Im Bereich der Geschäftsprozessautomatisierung ist dies ein vielversprechender Ansatz [Eg20]. Beispielsweise können Smart Contracts in den Lieferketten (Supply Chain) eingesetzt werden, um den Warenverkehr zwischen Parteien zu steuern und automatische Bestellungen sowie Zahlungen durchzuführen. Außerdem erwähnen Bhagavan, et. al. [Bh21], dass laut Gartner (Studie für Marktforschungsergebnisse und Analyse über die Entwicklung der IT) Unternehmen durch den Einsatz von Blockchains und Smart Contracts die Gesamtdatenqualität bis 2023 um 50% steigern werden.

5 Zusammenfassung und Ausblick

Die vorliegende Arbeit hat sich mit den Ansätzen zur Geschäftsprozessautomatisierung beschäftigt. Die Automatisierung von Geschäftsprozessen ist ein wichtiger Bestandteil eines jeden Unternehmens. Durch die Automatisierungen werden somit manuelle und wiederkehrende Aufgaben beseitigt. Solche Automatisierungen können das Auftreten von Fehlern und die Kosten, die bei einer manuellen Ausführung entstehen würden, verringern. Insgesamt wurden acht Ansätze zur Geschäftsprozessautomatisierung identifiziert, beschrieben, und die konkreten Technologien zu deren Umsetzung erläutert. Mit dieser Arbeit wurde somit ein Überblick über die Ansätze zur Automatisierung von Geschäftsprozessen geschaffen, welche für Unternehmen hilfreich sein könnten, um ihr Know-how über das Thema Geschäftsprozessautomatisierung zu erweitern. Zusammenfassend kann festgestellt werden, dass Unternehmen in der heutigen Zeit eine Bandbreite an Automatisierungsansätzen zur Auswahl stehen. Der Stand der Wissenschaft hat sich in den letzten Jahren stetig weiterentwickelt, wodurch zunehmend komplexere Ansätze entstehen. Die potenziellen Anwendungsbereiche werden damit erweitert und die Automatisierung dringt somit immer mehr in Arbeitsbereiche ein.
Hinsichtlich der gewählten Vorgehensweise muss angemerkt werden, dass im Rahmen dieser Arbeit nur eine begrenzte Anzahl von Literaturdatenbanken untersucht wurde und daher einzelne Ansätze, Technologien oder Merkmale ggf. nicht abgebildet wurden. Für zukünftige Arbeiten wäre es daher interessant, weitere Ansätze zu analysieren. Zusätzlich könnte untersucht werden, welche Automatisierungsansätze sich für welche Art von Unternehmen rentieren. Die Einführung des passenden Automatisierungsansatzes für die richtigen Geschäftsprozesse eines Unternehmens könnte die Effizienz der Prozesse steigern und dabei signifikant Zeit- und Kosteneinsparungen ermöglichen. Darüber hinaus könnte geforscht werden, welchen Einfluss die Prozessautomatisierung hinsichtlich der künstlichen Intelligenz auf den Einzelnen und die Gesellschaft hat. Immer mehr trägt die KI dazu bei, Entscheidungen zu treffen und übernimmt somit die Rolle des Menschen in gewissen Arbeitsbereichen. Diese Aussage bildet somit die Hypothese: Durch den Einsatz von KI wird die Automatisierungsquote in den nächsten Jahren weiterhin steigen.

Literatur

[Sc18] Schwarz, L., Neumann, T., Teich, T.: Grundlagen des Business Process Managements. Geschäftsprozesse praxisorientiert modellieren: Handbuch zur Reduzierung der Komplexität, pp. 21-42. Springer Berlin Heidelberg, Berlin, Heidelberg (2018)
[Ch17] Chakraborti, T., Isahagian, V., Khalaf, R., Khazaeni, Y., Muthusamy, V., Rizk, Y., Unuvar, M.: From Robotic Process Automation to Intelligent Process Automation. pp. 215-228. Springer International Publishing, 2017
[Me17] Mertens, P., Bodendorf, F., König, W., Schumann, M., Hess, T., Buxmann, P.: Integrierte Anwendungssysteme im Unternehmen. Grundzüge der Wirtschaftsinformatik, pp. 63-129. Springer Berlin Heidelberg, Berlin, Heidelberg (2017)

[Sh19] Shubha, R.: Impact of Automation and Artificial Intelligence. International Journal of Human Computing Studies 1, 14-17 (2019)
[Du21] Dumas, M., La Rosa, M., Mendling, J., Reijers, H.A.: Prozessorientierte InformationssystemeInformationssysteme, prozessorientierte. Grundlagen des Geschäftsprozessmanagements: übersetzt von Thomas Grisold, Steven Groß, Jan Mendling, Bastian Wurm, pp. 399-432. Springer Berlin Heidelberg, Berlin, Heidelberg (2021)
[Gae20] Gärtner, C.: Grundlagen: KI, ML, DL, RPA und Co. Smart HRM: Digitale Tools für die Personalarbeit, pp. 17-50. Springer Fachmedien Wiesbaden, Wiesbaden (2020)
[Fe06] Fettke, P.: State-of-the-Art des State-of-the-Art. WIRTSCHAFTSINFORMATIK 48, 257 (2006)
[WW02] Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly xiii-xxiii (2002)
[Ga20] Gadatsch, A.: Einführung in das Geschäftsprozessmanagement. In: Gadatsch, A. (ed.) Grundkurs Geschäftsprozess-Management: Analyse, Modellierung, Optimierung und Controlling von Prozessen, pp. 1-18. Springer Fachmedien Wiesbaden, Wiesbaden (2020)
[Ra20] Ralf, L., Koschmider, A.: 9 Geschäftsprozessmanagementsysteme und Robotic Process Automation. In: Ralf, L., Agnes, K., Dirk, F. (eds.) Prozessmanagement und Process-Mining: Grundlagen, pp. 153-164. De Gruyter Oldenbourg (2020)
[Cz18] Czarnecki, C., Auth, G.: Prozessdigitalisierung durch Robotic Process Automation. In: Barton, T., Müller, C., Seel, C. (eds.) Digitalisierung in Unternehmen: Von den theoretischen Ansätzen zur praktischen Umsetzung, pp. 113-131. Springer Fachmedien Wiesbaden, Wiesbaden (2018)
[Fe21] Fernando, E., Kriswanto, S.P.H., Wifasari, S.: Enterprise Resource Planning Systems: The Business Backbone. 2021 The 5th International Conference on E-Commerce, E-Business and E-Government, pp. 43–48. Association for Computing Machinery Meiryani (2021)
[Du18] Durand, C.: Internationalizing Mainframe Applications through Screen Scraping. pp. 228-235. Springer Berlin Heidelberg, 2018
[Me18] Mendling, J., Decker, G., Hull, R., Reijers, H., Weber, I.: How do Machine Learning, Robotic Process Automation, and Blockchains Affect the Human Factor in Business Process Management? Communications of the Association for Information Systems 43, 297-320 (2018)
[Ri21] Rizk, Y., Chakraborti, T., Isahagian, V., Khazaeni, Y.: 8 Towards end-to-end business process automation: RPA composition and orchestration. In: Christian, C., Peter, F. (eds.) Robotic Process Automation: Management, Technology, Applications, pp. 155-168. De Gruyter Oldenbourg (2021)
[Ji17] Jiménez-Ramírez, A.: Humans, Processes and Robots: A Journey to Hyperautomation. pp. 3-6. Springer International Publishing, 2017
[Ta20] Taulli, T.: RPA Foundations. The Robotic Process Automation Handbook: A Guide to Implementing RPA Systems, pp. 1-25. Apress, Berkeley, CA (2020)
[La21] Langmann, C., Turi, D.: Grundlagen zu Robotic Process Automation (RPA). Robotic Process Automation (RPA) - Digitalisierung und Automatisierung von Prozessen : Voraussetzungen, Funktionsweise und Implementierung am Beispiel des Controllings und Rechnungswesens, pp. 5-18. Springer Fachmedien Wiesbaden, Wiesbaden (2021)
[Ma20] Matthews, P., Greenspan, S.: Robotic Process Automation. Automation and Collaborative Robotics: A Guide to the Future of Work, pp. 71-107. Apress, Berkeley, CA (2020)
[Sm19] Smeets, M., Erhard, R., Kaußler, T.: Robotic Process Automation – Hintergründe und Einführung. Robotic Process Automation (RPA) in der Finanzwirtschaft: Technologie – Implementierung – Erfolgsfaktoren für Entscheider und Anwender, pp. 7-36. Springer Fachmedien Wiesbaden, Wiesbaden (2019)

[Ha21] Haleem, A., Javaid, M., Singh, R.P., Rab, S., Suman, R.: Hyperautomation for the enhancement of automation in industries. Sensors International 2, 100124 (2021)
[He21] Herrmann, M., Horstmann, M., Kaper, N.: Hyperautomation. In: Detscher, S. (ed.) Digitales Management und Marketing: So nutzen Unternehmen die Marktchancen der Digitalisierung, pp. 505-523. Springer Fachmedien Wiesbaden, Wiesbaden (2021)
[Ha20] Hammer, M., Safar, M., Niewel, T., Petzka, P., Liese, P., Rahnavard, S.: Digitale Automatisierung. Digitale Welt 4, 76-87 (2020)
[Kh21] Khan, S.N., Loukil, F., Ghedira-Guegan, C., Benkhelifa, E., Bani-Hani, A.: Blockchain smart contracts: Applications, challenges, and future trends. Peer-to-Peer Networking and Applications 14, 2901-2925 (2021)
[Ma19] Mammadzada, K., Iqbal, M., Milani, F., García-Bañuelos, L., Matulevičius, R.: Blockchain Oracles: A Framework for Blockchain-Based Applications. pp. 19-34. Springer International Publishing, 2019
[Eg20] Eggers, J., Hein, A., Weking, J., Böhm, M., Krcmar, H.: Process Automation on the Blockchain: An Exploratory Case Study on Smart Contracts (2020)
[Bh21] Bhagavan, S., Rao, P., Njilla, L.: A Primer on Smart Contracts and Blockchains for Smart Cities. In: Augusto, J.C. (ed.) Handbook of Smart Cities, pp. 1131-1161. Springer International Publishing, Cham (2021)

14. Workshop KI-basiertes Management und Optimierung komplexer Systeme / MOC 2022

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1655

The Proportional Constraint and Its Pruning – Continued[1]

Armin Wolf[2]

Abstract: Motivated by the necessity to model the adaptation of water levels in locks, a new variant of the *Proportional Constraint* is introduced in finite integer domain Constraint Programming using rounding-up (ceiling) instead of rounding. For its practical use in applications of finite domain Constraint Programming pruning rules are presented and their correctness is proven. Further, it is shown by examples that the number of iterations necessary to reach a fixed-point while pruning depends on the considered constraint instances. Importantly, fixed-point iteration always results in the strongest notion of bounds consistency which is proved, too. Furthermore, an alternative modelling of this constraint is presented. The run-times of the implementations of both alternatives are compared showing that the pruning rules introduced herein perform always better than the alternative approach on the chosen problem samples.

Keywords: bounds consistency; finite domain Constraint Programming; fixed-point iteration; Proportional Ceiling Constraint; pruning rules

1 Introduction

In [Wo18] we already introduced the *Proportional Constraint*. There, the necessity for this constraint was motivated by the need to model energy losses of batteries or the ratio between heating power and electric power of combined heat and power (CHP) systems in finite domain Constraint Programming.

Within the publicly funded project *SchleusenNOK40*[3] we are optimising the operation of sea locks, even the scheduling and allocation of ships in lock chambers in maritime traffic. This lock allocation and scheduling problem, especially the placement problem of the ships in the lock chambers can be considered as constraint-based resource scheduling problems, where Constraint Programming is the method of choice [BPN01]. One of the sub-problems to be solved is, among other things, the necessary adjustment of water levels depending on the difference between the water levels on both sides of the lock chambers and the speed for lowering or raising the water level, e. g. in [cm/min], which is only one aspect of this highly specialised problem.

Constraint Programming is well-suited to solve resource allocation and scheduling problems (cf. [BPN01]). However due to its nature, finite *integer* domain Constraint Programming

[1] The presented work is funded by the German Federal Ministry for Digital and Transport within the project "SchleusenNOK40" (BMDV, funding number: 19F2119B).

[2] Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany, armin.wolf@fokus.fraunhofer.de

[3] cf. https://www.fokus.fraunhofer.de/de/viscom/project/schleusennok40

doi:10.18420/inf2022_142

only supports integer variables such that linear equations are considered as *diophantine equations*, i. e. only the integer solutions are sought. However, this is not suitable for any modelling the required time for adjusting the water levels in lock chambers as the following example shows:

Example 1 *Let the gauge level at the entrance of a lock chamber be 457 [cm] and the gauge level at the exit side of the same lock chamber be 328 [cm]. The speed for lowering the water level from 457 [cm] to 328 [cm] be 21 [cm/min]. For a robust scheduling of lock operation we interested in*

$$\left\lceil \frac{457[cm] - 328[cm]}{21[cm/min]} \right\rceil = \lceil 6.1428... \rceil \ [min] = 7 \ [min] \ , \tag{1}$$

i. e. a 'conservative' rounding of the computed time using the ceiling function $\lceil . \rceil$.

It should be noted the gauge levels in our project are tide-dependent, i. e. they directly depend on the closing and opening times of the lock chambers such that in general we focus on the modelling of

$$\left\lceil \frac{|X - Y|}{s} \right\rceil \ , \tag{2}$$

where X and Y are finite domain integer constraint variables and s is a scalar factor. Furthermore the values of the variables X and Y are depending on some variable time points S and T such that X and S as well as Y and T are interrelated by *Element*[4] constraints mapping time points S and T to gauge levels X and Y.

The work is organised as follows: First we present some related work, then we define the *Proportional Ceiling Constraint* and according pruning rules. Further, the correctness of these rules is proven and it is shown that the number of iterations for reaching a fixed-point depend on the problem setting. However, fixed-point iteration leads to the strongest notion of bounds consistency. Then, an alternative modelling of the *Proportional Ceiling Constraint* based on weighted sums is presented and the run-times of these two approaches on sample problem classes are compared. Finally we conclude with some remarks on the implementation and the use of this constraint.

2 Related Work

Linear equations $Y = \alpha \cdot X$ where the variable Y *is proportional to* another variable X – where α is a scalar value – are special cases of *weighted sums*, i. e. $Y = \alpha_1 \cdot X_1 + \cdots + \alpha_n \cdot X_n (n > 0)$. *Weighted sum constraints* are already considered in [SS05]. Applying the pruning rules defined there on finite domain integer variables, the resulting consistency only ensures

[4] cf. https://sofdem.github.io/gccat/gccat/Celement.html

that there exist *real* solutions which is not adequate in the context of our *SchleusenNOK40* project. There, the variable values must be integral and rounding-up of non-integral values is required. Thus, we decided to further extend our object-oriented constraint solving library firstCS [Wo12] which already supports *Proportional Constraints* [Wo18] with an adequate *Proportional Ceiling Constraint* $Y = \alpha \cdot X$ for *finite domain integer variables*, where rounding is replaced by rounding-up.

3 The Proportional Ceiling Constraint

In application domains of finite domain Constraint Programming such as the optimised maritime lock scheduling there is a need to model the duration for adjusting the water level in a lock *proportional* to the speed of its change. Therefore and for other applications as well, we define the binary *Proportional Ceiling Constraint*:

Definition 1 (Proportional Ceiling Constraint) *Let $t > 0$ be a real value and A, B finite domain constraint variables having integer domains $\mathsf{dom}(A)$ respective $\mathsf{dom}(B)$. For convenience, let $\min(X) = \min(\mathsf{dom}(X))$ and $\max(X) = \max(\mathsf{dom}(X))$ for any domain variable X. The (binary)* Proportional Ceiling Constraint

$$\lceil t \cdot A \rceil = B \tag{3}$$

is satisfied, if for any value $a \in \mathsf{dom}(A)$ there is a value $b \in \mathsf{dom}(B)$ and if for any value $b \in \mathsf{dom}(B)$ there is a value $a \in \mathsf{dom}(A)$ such that $\lceil t \cdot a \rceil = b$ holds. Such value pairs (a, b) or labelings $\Theta = \{A \mapsto a, B \mapsto b\}$ satisfying *the constraint are called* solutions. *There, $\lceil . \rceil$ is the* rounding-up *(ceiling) function where $\lceil y \rceil$ is the smallest integer value greater than or equal to y for any real value y, i. e. $\lceil y \rceil \in \mathbb{N}$ and $\lceil y \rceil - 1 < y \leq \lceil y \rceil$.*

The definition of the *Proportional Ceiling Constraint* is *sound* in the sense that for any $t > 0$ and any integer value a there is another integer value b such that $\lceil t \cdot a \rceil = b$ holds: For any integer value a there is obviously $b = \lceil t \cdot a \rceil$. For any integer value b and $0 < t \leq 1$ there is always an integer value $a \in [(b - 0.5)/t, (b + 0.5)/t]$ satisfying the constraint, because it holds

$$t \leq 1 \Leftrightarrow t \leq (b + 0.5) - (b - 0.5) \Leftrightarrow 1 \leq \frac{b + 0.5}{t} - \frac{b - 0.5}{t} \tag{4}$$

However, for $t > 1$ this is not always the case: Let $b = 1$ and $t = 1.1$, then there is not any integer value a such that $1 = \lceil 1.1 \cdot a \rceil$ holds.

In order to implement and use such a constraint in a Constraint Programming system some *pruning rules* have to be defined, reducing the domains of the involved variables without losing any solutions and resulting in a fixed-point when iterated such that the pruned domains of the variables hopefully satisfy some notion of consistency. Our definition of some pruning rules for the Proportional Ceiling Constraint requires a special kind of an adapted „ceiling“ function mapping reals to integers. It is defined as follows:

Definition 2 *For any real value x let the function $\uparrow.\uparrow : \mathbb{R} \rightarrow \mathbb{Z}$ be defined by*

$$\uparrow x\uparrow = \begin{cases} x+1 & \text{if } x = \lceil x \rceil, \\ \lceil x \rceil & \text{otherwise.} \end{cases}$$

This definition of the function $\uparrow.\uparrow$ is sound in the sense that for any $x \in \mathbb{R}$ there is exactly one $y \in \mathbb{Z}$ such that $y = \uparrow x\uparrow$ holds.

For the defined Proportional Ceiling Constraint we propose the following *pruning rules*:

Definition 3 (Pruning Rules) *For any Proportional Ceiling Constraint*

$$\lceil t \cdot A \rceil \quad = \quad B$$

with $t > 0$ and finite domain constraint variables A and B having integer domains $\mathsf{dom}(A)$ respective $\mathsf{dom}(B)$ let

$$\mathsf{dom}^*(B) \quad = \quad \mathsf{dom}(B) \cap [\lceil t \cdot \min(A) \rceil, \lceil t \cdot \max(A) \rceil] \tag{5}$$

$$\mathsf{dom}^*(A) \quad = \quad \mathsf{dom}(A) \cap [\uparrow(\min^*(B) - 1)/t\uparrow, \lfloor \max^*(B)/t \rfloor] \tag{6}$$

be some pruning rules *– to be applied in the given order – where* $\min^*(B) = \min(\mathsf{dom}^*(B))$ *and* $\max^*(B) = \max(\mathsf{dom}^*(B))$ *for the sake of convenience. Further, let $\lfloor y \rfloor$ be the greatest integer value less than or equal to y for any real value y, i. e. $\lfloor y \rfloor \in \mathbb{N}$ and $\lfloor y \rfloor \leq y < \lfloor y \rfloor + 1$.*

These rules are potentially reducing the domains of A and B, i. e. $\mathsf{dom}^*(A) \subseteq \mathsf{dom}(A)$ and $\mathsf{dom}^*(B) \subseteq \mathsf{dom}(B)$ hold.

The indicator * in $\mathsf{dom}^*(A)$ respective in $\mathsf{dom}^*(B)$ is used to distinguish between the original and the updated domains of the variables A and B which will replace $\mathsf{dom}(A)$ respective $\mathsf{dom}(B)$ in any next iteration of these pruning rules.

Example 2 *Let the Proportional Ceiling Constraint $\lceil 2.1 \cdot A \rceil = B$ with $\mathsf{dom}(A) = \{0, 1, 2, 3\}$ and $\mathsf{dom}(B) = \{3\}$ be given. After applying the pruning rules defined in Definiton 3 it holds that $\mathsf{dom}^*(B) = \{3\} \cap [0, 7] = \{3\}$ and $\mathsf{dom}^*(A) = \{0, 1, 2, 3\} \cap [1, 1] = \{1\}$. This means that pruning determines the solution $\Theta = \{A \mapsto 1, B \mapsto 3\}$ correctly.*

Obviously, the question arises whether the pruning rules are *correct* or whether there are any integer solutions of the constraint which will be lost while pruning? – The following proposition answers this question:

Proposition 1 *Let $t > 0$ be a real value and A, B finite domain constraint variables having integer domains $\mathsf{dom}(A), \mathsf{dom}(B)$. If there is a value pair $(a, b) \in \mathsf{dom}(A) \times \mathsf{dom}(B)$ such that $b = \lceil t \cdot a \rceil$ holds, then $(a, b) \in \mathsf{dom}^*(A) \times \mathsf{dom}^*(B)$ will hold, too. In other words the propagation rules are correct, i. e. not any integer solution is lost while applying the pruning rules in Definiton 3.*

Proof 1 *Let $a \in \mathsf{dom}(A)$ and $b \in \mathsf{dom}(B)$ be any two integer values such that $\lceil t \cdot a \rceil = b$ holds. Due to the fact that $\min(A) \leq a \leq \max(A)$ holds and the function $f(x) = \lceil t \cdot x \rceil$ is monotonic, it holds $f(\min(A)) \leq f(a) \leq f(\max(A))$ and thus $b \in \mathsf{dom}^*(B)$ (cf. Equation (5)). According to the definition of $\lceil \_ \rceil$ it holds that:*

$$t \cdot a = \lceil t \cdot a \rceil - \varepsilon \quad \textit{with} \quad 0 \leq \varepsilon < 1 \tag{7}$$

Due to the facts that $b \in \mathsf{dom}^*(B)$ and $b = \lceil t \cdot a \rceil$ it holds that

$$\begin{aligned} \lceil (\min^*(B) - 1)/t \rceil &\leq \lceil (b-1)/t \rceil = \lceil (\lceil t \cdot a \rceil - 1)/t \rceil \\ &= \lceil (t \cdot a + \varepsilon - 1)/t \rceil = \lceil a - (1-\varepsilon)/t \rceil \\ &\leq a \quad \textit{because } (1-\varepsilon)/t > 0. \end{aligned} \tag{8}$$

Further, it holds that

$$\begin{aligned} \lfloor \max^*(B)/t \rfloor &\geq \lfloor b/t \rfloor = \lfloor \lceil t \cdot a \rceil / t \rfloor \\ &= \lfloor (t \cdot a + \varepsilon)/t \rfloor = \lfloor a + \varepsilon/t \rfloor \\ &\geq a \quad \textit{because } \varepsilon/t \geq 0. \end{aligned} \tag{9}$$

□

An iterated application of the pruning rules defined in Definition 3 on the constraint variables' domains either reduces these finite domains until they become empty or will not be further reduced. In any case, the iteration stops after a finite number of steps, such that

$$\mathsf{dom}^*(A) = \mathsf{dom}(A) \quad \text{and} \quad \mathsf{dom}^*(B) = \mathsf{dom}(B) \tag{10}$$

holds, i. e. a *fixed-point* is reached (cf. [Ap97]). Then any further applications of the pruning rules will not change the domains of the variables. However, how many iterations are necessary for reaching a fixed-point? – The following proposition answers this question:

Proposition 2 *The number of iterations of the pruning rules necessary to reach a fixed-point has no fixed upper bound: The number of iterations strongly depends on the constraint instance, in particular on the structure and on the size of the domains of the variables.*

Proof 2 (by a counter examples) *We assume the number of iterations is below an upper bound* $N \in \mathbb{N}$. *Now let* $t = 3.0$, $n \in \mathbb{N}$ *with* $n \gg N$, $\mathsf{dom}(A) = \{1, 2, 3, \ldots, n\}$ *and* $\mathsf{dom}(B) = \{2, 5, 8, \ldots, 3n - 1\}$ *be given. Then a fixed-point is reached after at least* $n - 1$ *iterations. The same holds for* $t = 0.3$ *an integer value* $n \gg N$, $\mathsf{dom}(A) = \{10, 20, 30, \ldots, 10 \cdot n\}$ *and* $\mathsf{dom}(B) = \{2, 5, 8, \ldots, 3n - 1\}$. *Then a fixed-point is reached after at least* $n - 1$ *iterations, too.*[5] □

Finally we show that after a fixed-point iteration of the pruning rules (cf. Def. 3) the domains of the variables of the *Proportional Ceiling Constraint* are *bounds consistent* in the strongest sense – cf. [Ch06] for a detailed analysis of different notions of *bounds consistency*. From there we adopted the following definition:

Definition 4 *A domain* D *is bounds*$(\mathcal{D})$ *consistent for a constraint* c *where* $\mathsf{vars}(c) = \{x_1, \ldots, x_n\}$, *if for each variable* x_i *with* $1 \leq i \leq n$ *and for each* $d_i \in \{min(x_i), \max(x_i)\}$ *there exist integers* d_j *with* $d_j \in \mathsf{dom}(x_j)$ *where* $1 \leq j \leq n$, $j \neq i$ *such that the labeling* $\Theta = \{x_1 \mapsto d_1, \ldots, x_n \mapsto d_n\}$ *is an integer solution of* c.

This definition considers n-ary constraints and thus binary constraints like the Proportional Ceiling Constraint as well.

Proposition 3 *Let a* Proportional Ceiling Constraint

$$\lceil t \cdot A \rceil \quad = \quad B$$

be given with $t > 0$ *and finite domain constraint variables* A *and* B *having integer domains* $\mathsf{dom}(A)$ *respective* $\mathsf{dom}(B)$. *Furthermore, it is assumed that the pruning rules (cf. Def. 3) are iterated until a fixed-point is reached, i. e.* $\mathsf{dom}^*(A) = \mathsf{dom}(A)$ *and* $\mathsf{dom}^*(B) = \mathsf{dom}(B)$ *hold. Then the domain* D *(cf. Def. 4) consisting of* $\mathsf{dom}(A)$ *and* $\mathsf{dom}(B)$ *is bounds*$(\mathcal{D})$ *consistent.*

Proof 3 (by case distinction) *Let* $a = \min(A)$. *Then, there is an integer value* b *such that* $b = \lceil t \cdot a \rceil$ *holds. Now we assume that* $b \notin \mathsf{dom}(B)$ *respective that* $b \neq \min(B)$. *It follows that* $b < \min(B)$ *due to the pruning rule (5) and the monotonicity of the ceiling function. Thus,* $\lceil t \cdot a \rceil \leq \min(B) - 1$ *holds and further* $t \cdot a \leq \min(B) - 1$. *This implies that* $a \leq (\min(B) - 1)/t$ *and finally* $a < \uparrow(\min(B) - 1)/t\uparrow$ *holds. This contradicts* $a = \min(A)$, *i. e. the assumption is wrong, it holds that* $b \in \mathsf{dom}(B)$ *respective that* $b = \min(B)$.

Let $a = \max(A)$. *Then, there is an integer value* b *such that* $b = \lceil t \cdot a \rceil$ *holds. Now we assume that* $b \notin \mathsf{dom}(B)$ *respective that* $b \neq \max(B)$. *It follows that* $b > \max(B)$ *due to*

[5] The detailled formal proof by induction over n is left to the interested reader.

the pruning rule (5) and the monotonicity of the ceiling function. Thus, $\lceil t \cdot a \rceil \geq \max(B) + 1$ *holds and further* $t \cdot a \geq \max(B) + 1$. *This implies that* $a \geq (\max(B) + 1)/t$ *and finally* $a > \lceil \max(B)/t \rceil$ *because* $1/t > 0$ *holds. This contradicts* $a = \max(A)$, *i. e. the assumption is wrong, it holds that* $b \in \mathsf{dom}(B)$ *respective that* $b = \max(B)$.

Let $b = \min(B)$. *We further distinguish two additional sub-cases:*

(a) We further suppose that there is an integer value a such that $b = \lceil t \cdot a \rceil \in \mathsf{dom}(B)$ *holds. Now, we assume that* $a \notin \mathsf{dom}(A)$ *respective that* $a \neq \min(A)$. *It follows that* $a < \min(A)$ *due to the pruning rule (6) and the monotonicity of the ceiling function. Thus,* $b < \lceil t \cdot \min(A) \rceil$ *holds implying that* $b < \min(B)$. *This contradicts the assumption, thus* $a \in \mathsf{dom}(A)$ *respective that* $a = \min(A)$.

(b) Now, we assume that for each integer value a it holds that $b \neq \lceil t \cdot a \rceil$ *even for* $a = \min(A)$. *Thus,* $b \neq \lceil t \cdot \min(A) \rceil$ *holds. According to Pruning Rule (5) it holds* $b > \lceil t \cdot \min(A)) \rceil$ *and thus* $b - 1 \geq t \cdot \min(A)$. *It follows that* $\lceil (b-1)/t \rceil > \min(A)$, *i. e.* $\min(A) \notin \mathsf{dom}(A)$. *This contradicts the assumption, i. e. there is an integer value a such that* $b = \lceil t \cdot a \rceil$ *holds. The case (b) never occurs.*

Let $b = \max(B)$. *Again, we further distinguish two additional sub-cases:*

(a) We further suppose that there is an integer value a such that $b = \lceil t \cdot a \rceil$ *holds. Now, we assume that* $a \notin \mathsf{dom}(A)$ *respective that* $a \neq \max(A)$. *It follows that* $a > \max(A) > \lfloor \max(B)/t \rfloor$ *and thus* $\max(A) > \max(B)/t$. *It holds* $t \cdot a > t \cdot \max(A) > \max(B)$ *and thus* $\lceil t \cdot a \rceil > \max(B)$, *i. e.* $b > \max(B)$ *contradicting the precondition, thus* $a \in \mathsf{dom}(A)$ *respective that* $a = \max(A)$.

(b) Now, we assume that for each integer value a it holds that $b \neq \lceil t \cdot a \rceil$ *even for* $a = \max(A)$. *Thus,* $b < \lceil t \cdot \max(A) \rceil$ *holds. Thus it holds that* $\max(B) < t \cdot \max(A)$ *because b is an integer value. It follows* $\lfloor \max(B)/t \rfloor < \max(A)$. *This contradicts the Pruning Rule (6). The assumption is wrong, i. e. there is an integer value a such that* $b = \lceil t \cdot a \rceil$ *holds. The case (b) never occurs.* □

4 Alternative Modelling

For any *rational* factor $t > 0$ within a Proportional Ceiling Constraint $\lceil t \cdot A \rceil = B$ it is possible to model this constraint equivalently on the basis of a *weighted sum constraints* which are well established in finite domain Constraint Programming ([SS05]):

Proposition 4 *Let $t = p/q$ where p and q are positive integer values. Then, for any two finite domain constraint variables A, B the constraints*

$$t \cdot A \leq B \wedge B < t \cdot A + 1 \quad \textit{respective} \quad 0 \leq q \cdot B - p \cdot A \wedge q \cdot B - p \cdot A < q \tag{11}$$

are equivalent to the Proportional Ceiling Constraint $\lceil t \cdot A \rceil = B$ in the sense that any solution $\{A \mapsto a, B \mapsto b\}$ is a solution of (11) and vice-versa.

Proof 4 *Let $\{A \mapsto a, B \mapsto b\}$ be any solution of the Proportional Ceiling Constraint, i. e. $b = \lceil t \cdot a \rceil$. We distinguish two sub-cases:*

1. *Let $q = 1$, i. e. t be an integer value. It holds that $q \cdot b - p \cdot a = \lceil p \cdot a \rceil - p \cdot a = 0$. Obviously, $\{A \mapsto a, B \mapsto b\}$ is a solution of (11).*

2. *Let $q > 1$, i. e. t be non-integral. It holds that $q \cdot b - p \cdot a = q \cdot \lceil p/q \cdot a \rceil - p \cdot a = q \cdot ((p/q \cdot a) + \varepsilon) - p \cdot a = q \cdot \varepsilon$ and thus $0 \leq q \cdot \varepsilon < q$ (cf. Equation (7)).*

Now, let $\{A \mapsto a, B \mapsto b\}$ be any solution of (11), i. e. $0 \leq q \cdot b - p \cdot a \leq p$ holds and thus $0 \leq b - p/q \cdot a < 1$. Consequently, $b = \lceil p/q \cdot a \rceil$ is the only integer value satisfying this condition and thus $\{A \mapsto a, B \mapsto b\}$ is also a solution of the Proportional Ceiling Constraint. □

5 Run-Time Comparison

We compared the Proportional Ceiling Constraint and its alternative modelling (cf. (11)) on the problem classes already presented in [Wo18]. Therefore we implemented the Proportional Ceiling Constraint and its the pruning rules defined in Definition 3 in our finite domain constraint solving library firstCS [Wo12] which already supports *weighted sum constraints*. Then we considered the following classes of problem instances

- A(n): $t = 0.3$, $\mathsf{dom}(A) = \{10, 20, 30, \ldots, 10n\}$, $\mathsf{dom}(B) = \{3, 5, 8, \ldots, 3n - 1\}$.
- B(n): $t = 0.997$, $\mathsf{dom}(A) = \{1, 2, 3, \ldots, n\}$, $\mathsf{dom}(B) = \{1, 2, 3, \ldots, n\}$.
- C(n): $t = 0.003$, $\mathsf{dom}(A) = \{1, 2, 3, \ldots, n\}$, $\mathsf{dom}(B) = \{1, 2, 3, \ldots, n\}$.

For problem class A (cf. Example 2) we perform modelling and initial pruning for both approaches resulting in one solution. For the problem classes B and C we perform modelling, initial pruning as well as additional prunings while searching for all solutions. The search strategy – standard labelling – chooses variable A then B and their values in increasing

Instance	PCC avg	PCC best	ALT avg	ALT best	ALT/PCC avg	ALT/PCC best
A(10000)	18.1	2.0	29.5	18.0	163.0 %	900.0 %
A(20000)	27.8	15.0	42.5	31.0	152.9 %	206.7 %
A(40000)	29.6	15.0	87.7	63.0	296.3 %	420.0 %
A(80000)	51.6	31.0	142.1	125.0	275.4 %	403.2 %
B(10000)	41.7	20.0	53.0	31.0	127.1 %	155.0 %
B(20000)	59.2	32.0	82.9	60.0	140.0 %	187.5 %
B(40000)	100.8	80.0	146.3	120.0	145.1 %	150.0 %
B(80000)	154.4	50.0	236.3	100.0	153.0 %	200.0 %
C(10000)	38.4	15.0	41.9	22.0	109.1 %	146.7 %
C(20000)	54.8	40.0	78.9	50.0	144.0 %	125.0 %
C(40000)	98.1	71.0	140.8	111.0	143.5 %	156.3 %
C(80000)	155.3	40.0	251.5	101.0	161.9 %	252.5 %

Tab. 1: Run-time Comparison of the Proportional Ceiling Constraint and its Alternative

order while performing chronological backtracking in cases of detected inconsistencies. Pruning means that the according pruning rules are applied until a fixed-point is reached.

Table 1 shows the results of our experiments on the *Proportional Ceiling Constraint* (PCC) and its alternative modelling (ALT) executed on a Windows computer with Windows 10 Pro (64 bit), Intel i7 CPU, 2.60 GHz, 12 GByte RAM. The computations for all problem instances were repeated 10 times. We compared best run-times and average run-times (in msec.) showing that the execution of the *Proportional Ceiling Constraint* is always faster as already reported in [Wo18] for the *Proportional Constraint* on the same problem instances.

6 Generalisation

Obviously, the Proportional Ceiling Constraint can be extended for an arbitrary factor $t \in \mathbb{R}$:

- For $t = 0$ the pruning rules are trivial:
  - If $0 \in \mathsf{dom}(B)$ holds, $\mathsf{dom}^*(B) = \{0\}$ and $\mathsf{dom}^*(A) = \mathsf{dom}(A)$ will hold.
  - If $0 \notin \mathsf{dom}(B)$ holds, the constraint cannot be satisfied, i. e. $\mathsf{dom}^*(B)$ as well as $\mathsf{dom}^*(A)$ will become empty.
- For $t < 0$ let $B = \lceil -t \cdot A' \rceil$ where $A' = -A$ such that $\min(A) = -\max(A')$ and $\max(A) = -\min(A')$. Thus it holds

$$\begin{aligned} \mathsf{dom}^*(B) &= \mathsf{dom}(B) \cap [\lceil -t \cdot \min(A') \rceil, \lceil -t \cdot max(A') \rceil] \\ \mathsf{dom}^*(A') &= \mathsf{dom}(A') \cap [\uparrow(\min^*(B) - 1)/-t\uparrow, \lfloor \max^*(B)/-t \rfloor] \end{aligned}$$

and thus

$$\mathsf{dom}^*(B) = \mathsf{dom}(B) \cap [\lceil t \cdot \max(A) \rceil, \lceil t \cdot min(A) \rceil] \quad (12)$$

$$\mathsf{dom}^*(A) = \mathsf{dom}(A) \cap [-\lfloor \max^*(B)/-t \rfloor, -\uparrow(\min^*(B) - 1)/-t\uparrow] \quad (13)$$

Example 3 *Let the* generalised Proportional Ceiling Constraint $\lceil -2.1 \cdot A \rceil = B$ *with* $\mathsf{dom}(A) = \{0, 1, 2, 3\}$ *and* $\mathsf{dom}(B) = \{-4\}$ *be given. After applying the adapted pruning rules (12) and (13) it holds that* $\mathsf{dom}^*(B) = \{-4\} \cap [-6, 0] = \{-4\}$ *and* $\mathsf{dom}^*(A) = \{0, 1, 2, 3\} \cap [2, 2] = \{2\}$*, i. e. pruning determines the only solution* $\Theta = \{A \mapsto 2, B \mapsto -4\}$ *of this constraint satisfaction problem correctly.*

7 Conclusion and Future Work

This work continues our work on *Proportional Constraints* (cf. [Wo18]) in another application area (locking of ships instead of energy management). Herein we defined the *Proportional Ceiling Constraint* on finite integer domain constraint variables and presented according pruning rules. We proved that pruning based on these rules is correct and results in the strongest notion of bounds consistency (cf. [Ch06]). The introduced *Proportional Ceiling Constraint* is implemented in our object-oriented constraint solving library firstCS [Wo12] and compared with an alternative approach based on *weighted sum constraints* already available in firstCS.

The *Proportional Ceiling Constraint* is used in the context of the *SchleusenNOK40* project together with some *Element* constraints to model a subproblem, i. e. the duration for adapting the water level in sea locks depending on the different tide-dependent water levels on both sides of such locks.

Future work will focus on the definition and implementation of efficient algorithms yielding *arc consistency* (cf. [De03]) for the binary *Proportional Constraint* and the *Proportional Ceiling Constraint*.

Bibliography

[Ap97] Apt, Krzysztof R.: From Chaotic Iteration to Constraint Propagation. In (Degano, Pierpaolo; Gorrieri, Roberto; Marchetti-Spaccamela, Alberto, eds): Automata, Languages and Programming, 24th International Colloquium, ICALP'97, Bologna, Italy, 7-11 July 1997, Proceedings. Lecture Notes in Computer Science 1256. Springer, pp. 36–55, 1997.

[BPN01] Baptiste, Philippe; Pape, Claude Le; Nuijten, Wim: Constraint-Based Scheduling – Applying Constraint Programming to Scheduling Problems. Springer Verlag, 2001.

[Ch06] Choi, Chiu Wo; Harvey, Warwick; Lee, Jimmy Ho-Man; Stuckey, Peter J: Finite Domain Bounds Consistency Revisited. In: Australian Conference on Artificial Intelligence. volume 4304. Springer, pp. 49–58, 2006.

[De03] Dechter, Rina; Cohen, David et al.: Constraint processing. Morgan Kaufmann, 2003.

[SS05] Schulte, Christian; Stuckey, Peter J: When Do Bounds and Domain Propagation Lead to the Same Search Space? ACM Transactions on Programming Languages and Systems (TOPLAS), 27(3):388–425, 2005.

[Wo12] Wolf, Armin: firstCS – New Aspects on Combining Constraint Programming with Object-Orientation in Java. KI - Künstliche Intelligenz, 26(1):55–60, 2012.

[Wo18] Wolf, Armin: The Proportional Constraint and Its Pruning. In (Seipel, Dietmar; Hanus, Michael; Abreu, Salvador, eds): Declarative Programming and Knowledge Management. Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 53–63, 2018.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1667

Analyzing Smart Services from a (Data-) Ecosystem Perspective: Utilizing Network Theory for a graph-based Software Tool in the Domain Smart Living

Henrik Kortum[1], Simon Hagen[1] , Janis Hühn[1], Oliver Thomas[1]

Abstract: There has long been a trend away from monolithic solutions toward integrated service systems that combine technical services and product functionalities across different manufacturers. This is especially true for the Smart Living domain, where interconnected products and services are used in one of the most private areas. However, there is a lack of tools to manage these complex systems. This work represents the second iteration of an ongoing research project in which we are developing an analysis and management tool in response to specific requirements of various stakeholder groups of the Smart Living domain. We interpret the services and products of the data ecosystem as nodes of a network graph, which we analyze using established methods of network theory, e.g., to identify weak points and bottlenecks in the service systems. The metrics provided help domain experts and managers from the domain to perform concrete tasks and provide a business benefit.

Keywords: Smart Service System, Data Ecosystem, Network Theory, Smart Living, Design Science

1 Introduction

The demand for highly integrated services is growing steadily. Digitalization can act as this integrator by bundling offerings that have so far often been considered reclusively. Thereby it is creating increasing opportunities to realize additional value for consumers and producers [Va15]. This is illustrated for examples in the domain of Smart Living, which can be seen as an advancement of Smart Home. By connecting and integrating currently individual products and services, like single sensors or actuators, it enables autonomous service systems. In the Smart Living domain, these offerings concern by definition one of the most private sectors: Everyone's private apartment. Thus, characteristics such as comfort and security [Eb21] as well as context sensitivity [Ba20] have a significant impact on the attractiveness of Smart Living services and are only made possible by the integration of different components. The use case of the Intelligent Gatekeeper [Re22] illustrates these challenges prominently: Tangible and intangible goods with partly different, identical or overlapping functions of different manufacturers must work seamlessly together to provide the expected functionality [Ha19]. Therefore, the proliferation of Smart Living applications is still very limited, often due to its complexity. Nevertheless, the growth potential due to the increasing demand for this market [SF20] is correspondingly high.

[1] Deutsches Forschungszentrum für Künstliche Intelligenz GmbH – Smart Enterprise Engineering, Parkstraße 40, 49080 Osnabrück, Germany, {henrik.kortum, simon.hagen, janis.huehn, oliver.thomas}@dfki.de

doi:10.18420/inf2022_143

In the scientific community, especially in Information Systems and several Engineering disciplines, the offerings are often referred to as "product-service systems" or, through the now data-driven focus, as Smart Service (Systems) [Ko22]. Due to their integrative nature, they often have to incorporate several disciplines and overcome the heterogeneity of stakeholders involved in development and operation [La16]. Since the start of research into service bundling, their representation has been of great interest and thus been examined often [BT13]. Based on the consideration of individual components, e.g. the process perspective through process modelling [Ka19] or physical components through CAD design [Kl19], overarching approaches are becoming increasingly important [HKT18]. Publications can already be found in the scientific context, ranging from linking individual components, as Schoormann et al. [Sc20] propose for processes and business models, to thoroughly (meta) models of product-service systems (e.g. [MBB18]). However, no approach has yet been widely adopted allowing automated production and maintenance of these models. In addition, the previously outlined interplay of different actors combining their services, products and data sources leads to the emergence of a data ecosystem. In ecosystems, the data shared between actors and their services are an important resource with great potential for economic value creation. The management of this resource and of the entire data ecosystem is therefore of great importance but has not been adequately pursued to date.

Our research interest is generated by this fact and was initially met with the development of a prototype for the German Digital Summit 2020[2] [Re21], which provides a tool to gather, persist and represent (technical) self-descriptions of services and their relations and thereby acts as a Service Registry. Based on feedback from practitioners, the idea arose to use the resulting digital representation of the ecosystem for analytic purposes, e.g. to identify most critical services on which many others rely on. We have therefore systematically imposed requirements for such functionality and used them to further develop our tool. This paper thus focuses on developing a concept and implementing a prototype for analyzing ecosystems according to digital representations of the elements using graphs. Data ecosystems can be understood as networks [AB02] whose entities are represented as nodes, which are linked to each other via edges in a variety of relationships, forming complex graph structures. To analyze these graph structures, we use metrics and analytic methods from network theory and adapt them to our business-oriented use case. Because the proposed analysis is performed in real-time, we chose to use labeled property graphs instead of the established Resource Description Framework for their increased performance [DYB00]. We call this tool 'Smart Service Analyzer'.

In order to describe our approach thoroughly the paper is structured as follows. In Section 2 the scientific methods used are outlined and set into context of the overall research framework, which this paper is the second iteration of. Section 3 thus briefly illustrates our previous work and is followed by the description of the requirement elicitation (Section 4). The transfer from requirements into a functional and technical concept is explained in Section 5 and followed by the description of the implementation. Lastly, we discuss our work and point our future work opportunities in Section 7.

[2] https://www.de.digital/DIGITAL/Redaktion/DE/Dossier/digital-gipfel-2020.html

2 Methodical Approach

As stated before this paper represents the second iteration of a design science research (DSR) [He04] prototype development within ForeSight[3], a research project in the domain of Smart Living. In a previous paper [Re21], requirements for the Service Registry were elicited and a prototype was implemented based on these requirements. In the present paper we describe the evaluation of this prototype, which at the same time represents the requirements elicitation for the next iteration stage and the current status of the ongoing development process. Our methodological research approach thus follows the Design Science Research methodology (DSRM) of Peffers et al. [Pe07]. In this paper we focus primarily on the first three steps suggested in DSRM: problem identification, conceptualization of the solution and improvement of the solution artifact. The subsequent steps of demonstration and evaluation are not part of this work. Our DSRM oriented approach is outlined in Fig. 1.

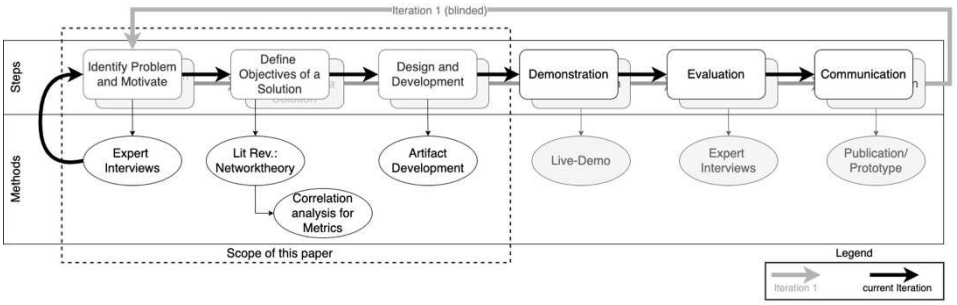


Fig. 1: Our DSRM [Pe07] oriented research approach

Since the evaluation of the prototype of the first iteration revealed the demand for an analysis function, we collected detailed requirements for such a function. These requirements enable us to incorporate domain knowledge into the development process, as suggested by [GH13]. For this purpose, we conducted focus groups with experts from different companies. All interviewed persons are experts in the sense of Bogner et al. [BLM09], hold a managerial position in their company and have several years of professional experience in the field of Smart Living. The results of the individual focus groups were subsequently consolidated in a workshop with all experts and are presented in section 4.

In a second step, we derive literature-based relevant methods and metrics for the realization of the analysis scenarios and the fulfilment of the previously identified requirements. Specifically, we draw on approaches from network theory, which we map to business problems in the Smart Living ecosystem. The selection of metrics is flanked by a correlation analysis, which helps us to identify metrics with identical meaning. Building on the insights gained here, we develop our solution artifact - the Smart Service Analyzer, a tool for analyzing data ecosystems.

[3] https://foresight-plattform.de

3 Representing the Smart Living Ecosystems

To enable innovative and secure service offerings, a transparent ecosystem maintaining interoperability and trust is needed. Within the beforementioned research project thus tools were implemented supporting these values and empowering companies utilizing the available services to build upon them or offer completely new services. One of these tools is the Service Registry, which was designed and developed as a catalog for collecting and presenting (digital) services. Therefor a standardized self-description for services was defined, which is a representation of individual services in the form of JSON-LD files. The definition is based on the W3C Web of Things standard[4] and was extended with Smart Living specific attributes. Services are stored in two levels of abstraction that build upon each other. 'Base Services' describe services in an abstract manner and focus their characteristics, while 'Service Instances' represent concrete instantiations of individual Base Services and contain, for example, information about specific endpoints. Base Services can be understood as building instructions for Service Instances. The Service Registry provides an interface to create and manage the services, both visually via a GUI and programmatically via an API. A similar approach was used for the standardized descriptions of Smart Home devices. These are stored in another catalog, the Thing Registry [Al21]. Together these tools provide information regarding elements in the ecosystem: physical goods, intangible services, datasets and their respective providers and thereby stakeholders.

4 Identifying Practical Needs and Potentials through Personas

The representation of entire ecosystems introduces a whole new level of complexity and thus barriers for wide adoption. To tackle these, focus groups with experts aimed at identifying ways to dissolve them and, most importantly, to gain insights on which benefits can arise from this integrated representation and analysis of an ecosystem, consisting e.g. of the self-descriptions in the Service Registry. Therefore, we chose an explorative and creative approach within the focus groups to discover aspects unthought of yet for different roles in the business context. As stated above, we firstly asked the participants to come up with personas for whom the tool can be beneficial, and secondly specify potentials in using it. In each focus group between one and three employees from the respective company shared their thoughts on requirements and potentials. In total we talked to 9 companies via video calls lasting about 45-60 minutes each. The participants cover all relevant areas of the Smart Living ecosystem, such as the housing industry, IoT and product manufacturers, ML developers, infrastructure providers, research institutes and service provider. During the focus groups, the respondents were asked to identify personas for whom an analysis function of the service graph is relevant and to derive concrete use cases and information needs for these personas.

[4] https://www.w3.org/WoT/

In Tab. 1 we present the personas and potentials we gathered. After we obtained the raw results, we grouped and aggregated them into the final list of most relevant aspects. In the first column the different roles are named. It was interesting to see, that the spread reached from technical (software development / data scientist) over business (product management / sales) to managerial (management) groups. Even though the tool can be of different importance for each group, it is promising that a wide target group exists. Assigned to each role are the stated potentials. One has to notice, that they partially have been associated with multiple roles by the experts. For simplicity we assigned it to the role it was most often mentioned with.

Business Role	**#**	**Potentials**
Product Manager	1.1	Identify gaps in the product portfolio
	1.2	Define product requirements
	1.3	„Supply chain“ monitoring
	1.4	Evaluate usage of offered services
	1.5	Calculate fair compensation
Software Developer	2.1	Risk assessment (e.g. failures on dependent services)
		2.1a: Risk for the Network
		2.1b: Risk for the Node
	2.2	Examining relationships
	2.3	Simplify troubleshooting
	2.4	Contact information
	2.5	Insights on technical feasibility
Data Scientist	3.1	Overview of present data sources
	3.2	Support in ensuring data quality
Sales	4.1	Identify potential customers
Management	5.1	Outline of portfolio
	5.2	Synopsis of business model (and opportunities)
	5.3	Review needs for certification
	5.4	Identify standard processes etc.
Operator / Ecosystem	6.1	Standardization
	6.2	Ensure privacy
	6.3	Ensure data sovereignty
	6.4	Identify / expand basis technology
	6.5	Overview over the ecosystem

Tab. 1: Potentials and thus functional requirements for the Smart Service Analyzer

5 Transforming Requirements into a technical Concept

5.1 Network Theory and Analysis in Data Ecosystems

In order to map the complex relationships between different entities of a data ecosystem (e.g. services, products, data, stakeholders), we use a graph representation and persistence of information in a graph database. In the same way that many complex systems can be described as networks of interacting entities [FS03], data ecosystems can also be

understood as complex networks [AB02] whose entities are represented as nodes, linked to each other via edges in a variety of relationships, e.g. hierarchical. In our case the ecosystem comprises various Smart Service Systems, which consist of different entities like Smart Services, data sources, products, or stakeholder thus, in the context of network theory, the ecosystem represents the entire network graph, individual Smart Services are subnetworks, and the previously mentioned entities represent individual nodes of various types. Fig. 2 illustrates the relations within the overall network.

This form of representing data ecosystems enables the use of statistical methods of network topology and dynamics, which provide conclusions about the characteristics of the system. In this paper, we are mainly interested in the relationships between Smart Services, which is why we limit our further analysis to the entity class "Services". In this way, various metrics [AS15] can be calculated, which allow statements about the robustness of the network to failures [AB02, BTL09] and provide starting points for its optimization. As different actors in data ecosystems develop and operate their services mainly independently, there is often no holistic view of the ecosystem as a whole. Although the services are dependent on each other, too little attention has been paid to this fact [Ko22]. As a result, critical dependencies on individual services can arise, so that their failure leads to a cascade of further service failures that affect sub-networks or even the entire ecosystem. From an economic perspective, such a failure is associated with high costs for the affected actors, which is why measures must be taken to avoid it. In addition to these relatively simple statistical metrics, more complex techniques, such as unsupervised machine learning methods, can also be used to identify patterns in the network structures of data ecosystems [Lo06] and to detect anomalies [ATK15].

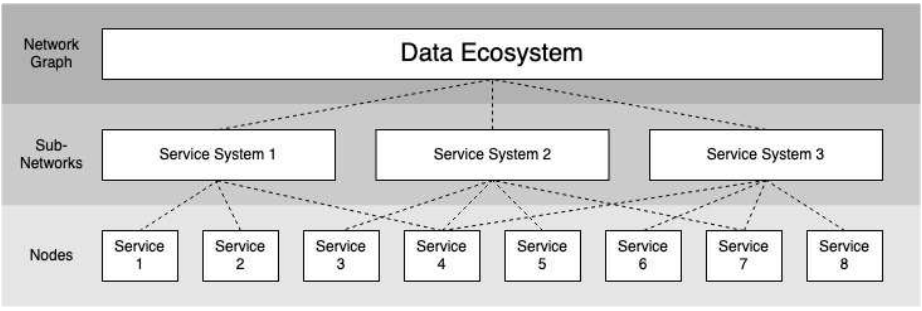


Fig. 2: Nodes and subnetworks in data ecosystems

5.2 Used Metrics and Methods

Metrics can be calculated at multiple levels of a data ecosystem. Firstly, for the entire data ecosystem, secondly, for individual, self-contained Smart Service Systems within the ecosystem and thirdly, at the level of individual services. In the literature of graph theory, there are a multitude of metrics that allow statements about the graph as a whole, subgraphs and individual nodes. To identify suitable metrics for meeting the business requirements identified in section 4, the relevant literature in this area was searched. Many

of the identified metrics have a strong overlap in content. Harmonic centrality, for example, is a variant of closeness centrality which solves the problem of dealing with disconnected graphs. Both metrics aim to determine the centrality of nodes in the network and can be used to identify services that are critical for the ecosystem. To avoid redundancies within the Analyzer and to counteract a cognitive overload of its users, we performed a correlation analysis on a test data set that served as a basis for decision-making for the selection of KPIs for the Smart Service Analyzer. Since the amount of data currently available from the Smart Living data ecosystem is not sufficient to make a significant statement about the correlation of metrics, we resorted to a sample of the Python package repository Python Package Index (PyPI). The sample consists of 2,179 packages and their relationships to each other.

The packages (services) are connected as nodes by their dependencies via edges and thus form a graph that can be interpreted in the same way as our Smart Living graph. Metrics with a Pearson correlation coefficient of more than 0.9 or less than -0.9 were examined for overlap in content, and in the case of a technically identical interpretation for the Smart Living Data Ecosystem, one of the metrics was removed. The results of the correlation analysis are shown in Fig. 3.

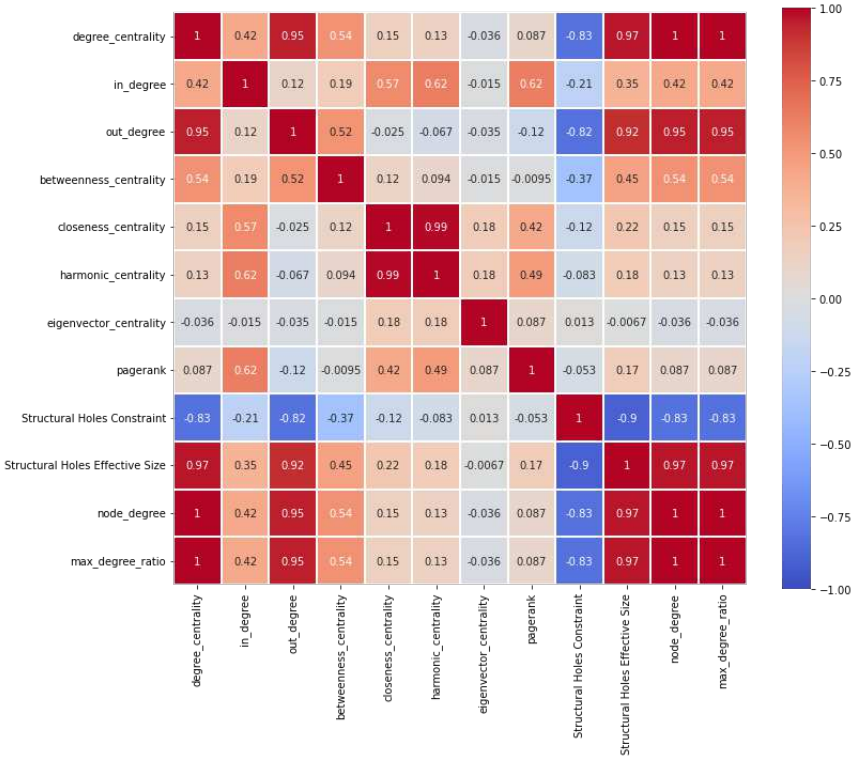


Fig. 3: Correlation matrix of tested metrics

The remaining metrics were evaluated by researchers with expertise in Smart Living and data ecosystems to obtain a domain specific interpretation of the metrics. Subsequently, a workshop was conducted to map the individual metrics to the requirements elicited through the focus groups. The results of this process are summarized in Tab. 2.

Req.	Metric	Description	Domain Specific Interpretation for the Smart Living Data Ecosystem
		Dimension: Services/Nodes	
1.2, 2.1b, 2.2	Indegree	Number of incoming edges of a node.	Number of services consumed by a service.
1.4, 2.1a, 2.2, 4.1, 5.2, 6.2, 6.3	Outdegree	Number of outgoing edges of a node.	Number of services that consume a service.
1.4, 1.5, 2.1a, 2.1b	(Normalized) Degree Centrality	Measures the number of direct relationships a node has, normalized by the maximum possible degree.	Allows information about how strongly interconnected a service is in relation to the size of the ecosystem.
1.4, 1.5, 2.1a, 5.2, 6.4	Betweenness centrality	Measures sum of the fraction of all-pairs shortest paths passing through a node [Br08].	Indicates how many shortest paths exist through the ecosystem via a given service. Enables the identification of important bridge services that connect other services of the service system.
1.5, 2.1a, 2.1b	PageRank	The PageRank is a variant of the Eigenvector metric, also taking link direction and weight into account. It mainly uses the indegree to estimate the influence level of nodes, thus applies to directed graphs [Pa99].	Allows conclusions to be drawn about the importance of a service to the data ecosystem. In addition to the number of services that Consume a given service, the number of services that consume the consuming services is also considered.
1.4, 1.5, 2.1a	Closeness centrality (Wasserman and Faust)	Calculates the inverse of the average shortest path between a node and all other nodes. Wasserman and Faust proposed an improved formula for usage with multi component graphs [WF94].	Allows conclusions to be drawn about how central and thus how important a service is for an isolated service system. The metric is suitable for our case because it works for data ecosystems in which not all nodes are connected, either directly or indirectly.

Dimension: Sub-Networks/Services Systems, Network Graph/Data Ecosystem			
1.1, 1.4, 2.2, 5.2, 6.2, 6.3	Structural Holes (Constraint)	Measures how strongly a given node is connected to nodes that are themselves connected to its neighbors [Bu04].	The occurrence of structural holes indicates that the connections between services in a system are unbalanced. This means that there are services with exclusive connections that link individual service systems in an ecosystem and whose failure could be critical for the data ecosystem as a whole.
2.1a	Max Degree Ratio	The Max Degree Ratio is the maximum degree of any node in the graph normalized by the highest possible degree [Il07].	Allows statements about whether there is at least one service node in the ecosystem with very high impact/damage potential in case of failure.
2.1a	Density	The Density of a graph is measured on a 0-1 scale where 1 is equivalent to a graph where every possible edge between two nodes exists and 0 is equivalent to a graph without any edges.	Represents the proportion of possible relationships between services in the data ecosystem that actually exist. A high value indicates strong dependencies of the ecosystem on individual services and thus the existence of individual points of failure.
Addition-ally: 1.3, 2.3, 2.4,	Exploratory analysis of the visualized network graph	-	Visual explorative analysis of the data ecosystem helps stakeholders get an overview of the services, the stakeholders connected to them, and can help with troubleshooting.

Tab. 2: Relevant metrics for the node and graph dimensions

6 Implementation

Transforming all the required information into a form suitable for graph analysis requires various technologies and technical artifacts, which will be described in further detail in the following paragraphs and in the sequence diagram (See Fig. 4). The existing service descriptions, which were previously stored exclusively in the relational database of the Service Registry, are converted into a format that represents the services and their stakeholders as nodes and their dependencies as edges. Thus, these service and stakeholder nodes and their edges can be transferred from the Service Registry into a graph database, for which we chose Neo4J [Ne22]. Nodes and relationships representing Smart Things (sensors and actuators) and their stakeholders are stored in the graph database as well. Based on the service description, a service may depend on a certain number of entities. The visualization of the data ecosystem is based on the nodes and edges stored in the graph

database and the Javascript library D3.js [BOH11]. The information of all nodes and edges are loaded from the Neo4j graph database into the Service Registry.

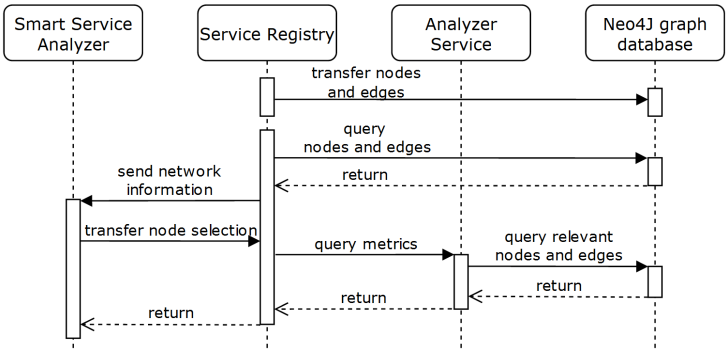


Fig. 4: Sequence diagram showing the Smart Service Analyzer and related technologies

Then these nodes and edges are systematically traversed and transformed using an algorithm to make them readable by D3.js. The Analyzer Service is a REST-full Python service that connects to the Neo4J database to retrieve the data for analysis. The service uses the Numpy and Pandas libraries for data manipulation and internal storage. Most of the network analysis is done with the open source library networkX [HSS00]. Upon receiving a generic request or a request containing a specific node ID the necessary information for the calculation of the metrics is loaded from the Neo4J graph database. While stakeholders and things are already represented in the graph database, they will not influence the analysis of the ecosystem as we solely focus on the service – service relationships in this work. Based on this information, the metrics are calculated and then sent back to the Service Registry via REST. This response is dynamically loaded into the web interface of the Smart Service Analyzer. Fig. 5 shows a screenshot of the Analyzer UI, illustrating how the data ecosystem graph and the metrics are displayed to users.

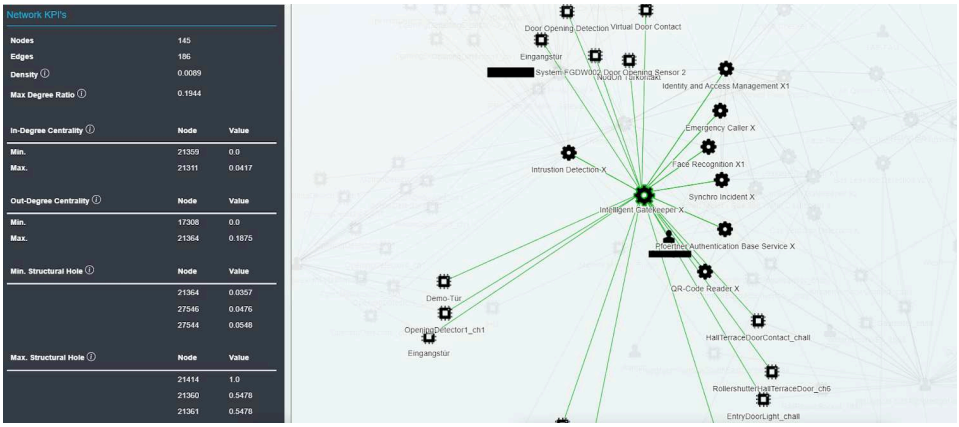


Fig. 5: Screenshot of the interface of the Smart Service Analyzer

7 Discussion, Limitations and Future Work

This paper presents a graph theory-based approach to analyze data ecosystems and a software tool demonstrating the approach, exemplarily in the domain Smart Living. Our proposition marks the second iteration of a design science research (DSR) cycle and thereby builds upon previous research insights and the corresponding prototype. The requirements were derived from focus groups and an aggregating workshop and mapped onto different established methods from graph analysis. It can take the different components data ecosystem currently encompass into account and thereby support companies offering highly integrated service offerings. By providing functionalities like planning, documenting, and analyzing services with respect to different metrics, our approach and thus the Service Registry and the Smart Service Analyzer provide a toolset for data ecosystems development.

However, in this paper we focus on the analysis of the relationships between services solely to demonstrate the approach. In future work it can be extended in several areas, e.g. including other ecosystem relevant entities, e.g., data sources, actors, or products as new node types. The introduction of new node classes would also result in new edge types to represent the relationships between the different nodes, giving the graph the additional property of being multi-relational. This can, for example, be integrated into the software with a toggle switching between the analysis of specific or all node types and edges. Also, the elements can be represented in greater detail (with more information). In addition, the set of metrics we have chosen for the Analyzer is only a snapshot covering the fulfilment of the collected requirements. In the course of time and with growing numbers of use cases, the requirements will change, and further metrics should be added to the Analyzer.

The field of machine learning-based analysis methods was only briefly touched upon in this paper. In addition to using the previously described metrics, more complex machine learning and clustering methods can be used to analyze network graphs, especially in the context of data mining to uncover previously unknown structures and relationships between services. In our further research work we plan to transform the network graphs of the data ecosystem into numerical feature vectors using methods like Node2Vec [GL16] or Graph2Vec [Na17], in order to further analyze them with different machine learning methods, such as DBSCAN or different classifiers.

An automatically classification or clustering of services can also be utilized for new forms of revenue allocation. It can be dynamically measured and e.g. oriented at the importance of an offering for the overall ecosystem to achieve a fair compensation.

One has to note that our tool only provides numeric metrics which support companies in the decision process. They do not contain qualitative solutions to overcome certain problems and thus implicitly enable the propositions defined in Tab 1. Therefore, methods and skills need to be examined, which empower people and companies to build upon the analysis results. A starting point for this is provided by the work of [KRT22] in which skill requirements for AI domains were examined based on the analysis of job advertisements.

A big advantage of our approach lies in its broad applicability for all data ecosystems, assuming that their entities are represented in a corresponding graph structure. This also results in the potential for cross-domain use, within which service systems between

different domains are analyzed. Here, we see added value for current initiatives to create transparent, interoperable ecosystems (e.g. GAIA-X) and would like to encourage practitioners and scientists to continue our work and use the Smart Service Analyzer.

8 Acknowledgements

We would like to thank our colleague Jonas Rebstadt for his support in developing the prototype and identifying the requirements. This research was conducted in the scope of the research project ForeSight, which is funded by the German Federal Ministry for Economic Affairs and Energy (BMWK). We would like to thank them for their support.

Bibliography

[Al21] Alberternst, S.; Anisimov, A.; Antakli, A.; Duppe, B.; Hoffmann, H.; Meiser, M.; Muaz, M.; Spieldenner, D.; Zinnikus, I.: Orchestrating Heterogeneous Devices and AI Services as Virtual Sensors for Secure Cloud-Based IoT Applications. In: Sensors vol. 21. MDPI, pp. 1-25, 2021.

[ATK15] Akoglu, L.; Tong, H.; Koutra, D.: Graph based anomaly detection and description: a survey. In: Data mining and knowledge discovery vol. 29. Springer, pp. 626–688, 2015.

[AB02] Albert, R.; Barabási, A.: Statistical mechanics of complex networks. In: Reviews of modern physics vol. 74. American Physical Society, pp. 47-97, 2002.

[AS15] Alenazi, M. J. F.; Sterbenz, J. P. G.: Comprehensive comparison and accuracy of graph metrics in predicting network resilience. In: 2015 11th International Conference on the Design of Reliable Communication Networks (DRCN). IEEE, pp. 157–164, 2015.

[Ba20] Bauer, J.; Hechtel, M.; Holzwarth, M.; Sessner, J.; Franke, J.; Wittenberg, T.; Waldhör, K.; Schena, A.; Falanga, A.; Ristok, B.; Böhle, F.: A Concept for Context Awareness in Smart Environments. In: Current Directions in Biomedical Engineering vol. 6. De Gruyter, pp. 380–383, 2020.

[BTL09] Bigdeli, A.; Tizghadam, A.; Leon-Garcia, A.: Comparison of network criticality, algebraic connectivity, and other graph metrics. In: Proceedings of the 1st Annual Workshop on Simplifying Complex Network for Practitioners. Association for Computing Machinery, pp. 1–6, 2009.

[BLM09] Bogner, A.; Littig, B.; Menz, W.: Introduction: Expert interviews—An introduction to a new methodological debate. In (Bogner, A., Littig, B., Menz, W, eds.): Interviewing experts. Palgrave Macmillan, London, pp. 1–13, 2009.

[BOH11] Bostock, M.; Ogievetsky, V.; Heer, J.: D3 Data-Driven Documents. In: IEEE Transactions on Visualization and Computer Graphics vol. 17. IEEE, pp. 2301–2309, 2011.

[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 vol. 51. Elsevier, pp. 245–260, 2013.

[Br08] Brandes, U.: On variants of shortest-path betweenness centrality and their generic computation. In: Social Networks vol. 30. Elsevier, pp. 136–145, 2008.

[Bu04] Burt, R. S.: Structural Holes and Good Ideas. In: American Journal of Sociology vol. 110. The University of Chicago Press, pp. 349–399, 2004.

[DYB00] Donkers, A. J. A.; Yang, D.; Baken, N.: Linked Data for Smart Homes: Comparing RDF and Labeled Property Graphs. In: LDAC2020—8th Linked Data in Architecture and Construction Workshop. Pp. 23-36, 2020.

[Eb21] Ebbers, F.; Zibuschka, J.; Zimmermann, C.; Hinz, O.: User preferences for privacy features in digital assistants. In: Electronic Markets vol. 31, pp. 411–426, 2021.

[FS03] Ferrer Cancho, R.; Solé Vicente, R.: Optimization in complex networks. In (Pastor-Satorras, R.; Rubi, M.; Diaz-Guilera, A, eds.): Statistical Mechanics of Complex Networks vol. 625. Springer, pp. 114–126, 2003.

[GH13] Gregor, S.; Hevner, A.: Positioning and Presenting Design Science Research for Maximum Impact. In: MIS Quarterly vol. 37. JSTOR, pp. 337–356, 2013.

[GL16] Grover, A.; Leskovec, J.: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. Association for Computing Machinery, pp. 855–864, 2016.

[HKT18] Hagen, S.; Kammler, F.; Thomas, O.: Adapting Product-Service System methods for the digital era – requirements for Smart PSS engineering. In: Customization 4.0. Springer Proceedings in Business and Economics. Springer, Cham, 2018.

[Ha19] Hagen, S.; Brinker, J.; Gembarski, P.C.; Lachmayer, R.; Thomas, O.: Integration von Smarten Produkten und Dienstleistungen im IoT-Zeitalter: Ein Graph-basierter Entwicklungsansatz. In: HMD, vol. 56. Pp. 1220–1232, 2022.

[HSS00] Hagberg, A.; Schult, D.; Swart, P.: NetworkX

[He04] Hevner, A. R.; March, S. T.; Park, J.; Ram, S.: Design Science in Information Systems Research. In: Management Information Systems Quarterly vol. 28. Management Information Systems Research Center, University of Minnesota, pp. 75-105, 2004.

[Il07] Iliofotou, M.; Pappu, P.; Faloutsos, M.; Mitzenmacher, M.; Singh, S.; Varghese, G.: Network monitoring using traffic dispersion graphs (tdgs). In: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement. Association for Computing Machinery, pp. 315-320, 2007.

[Kl19] Kloock-Schreiber, D.; Domarkas, L.; Gembarski, P.; Lachmayer, R.: Enrichment of Geometric CAD Models for Service Configuration. In: CEUR Workshop Proceedings 2467. Aachen, pp. 22-29, 2019.

[Ka19] Kammler, F.; Hagen, S.; Brinker, J.; Thomas, O.: LEVERAGING THE VALUE OF DATA- DRIVEN SERVICE SYSTEMS IN MANUFACTURING: A GRAPH-BASED APPROACH. In: Proceedings of the 27th European Conference on Information Systems (ECIS). Stockholm & Uppsala, pp. 1-15, 2019.

[KRT22] Kortum, H.; Rebstadt, J.; Thomas, O.: Dissection of AI Job Advertisements: A Text

Mining-based Analysis of Employee Skills in the Disciplines Computer Vision and Natural Language Processing. In: Proceedings of the 55th Hawaii International Conference on System Sciences. Pp. 5211-5220, 2022.

[Ko22] Kortum, H.; Rebstadt, J.; Hagen, S.; Thomas, O.: Integrating Data and Service Lifecycle for Smart Service Systems Engineering: Compilation of a Lifecycle Model for the Data Ecosystem of Smart Living. In: Proceedings of the 55th Hawaii International Conference on System Sciences. Pp. 2017-2026, 2022.

[La16] Larson, R. C.: Smart Service Systems: Bridging the Silos. In: Service Science 8(4). INFORMS, pp. 359–367, 2016.

[Lo06] Long, B.; Wu, X.; Zhang, Z.; Yu, P. S.: Unsupervised learning on k-partite graphs. In: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Association for Computing Machinery, pp. 317–326, 2006.

[MBB18] Maleki, E.; Belkadi, F.; Bernard, A.: A Meta-model for Product-Service System based on Systems Engineering approach. In: Procedia CIRP vol. 73. Elsevier, pp. 39–44, 2018.

[Na17] Narayanan, A.; Chandramohan, M.; Venkatesan, R.; Chen, L.; Liu, Y.; Jaiswal, S.: graph2vec: Learning distributed representations of graphs. In: arXiv preprint arXiv:1707.05005 (2017)

[Ne22] Neo4j Graph Data Platform – The Leader in Graph Databases, https://neo4j.com/, accessed: 01/02/2022.

[Pa99] Page, L.; Brin, S.; Motwani, R.; Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. URL http://ilpubs.stanford.edu:8090/422/. (2022-03-02)

[Pe07] Peffers, K.; Tuunanen, T.; Rothenberger, M. A.; Chatterjee, S.: A Design Science Research Methodology for Information Systems Research. In: Journal of Management Information Systems vol. 24. Routledge, pp. 45–77, 2007.

[Re22] Rebstadt, J.; Kortum, H.; Gravemeier, L.S.; Eberhardt, B.; Thomas, O.: Non-Discrimination-by-Design: Handlungsempfehlungen für die Entwicklung von vertrauenswürdigen KI-Services. In: HMD, vol. 59. Pp. 495–511, 2022.

[Re21] Rebstadt, J.; Kortum, H.; Hagen, S.; Thomas, O.: Towards a transparency-oriented and integrating Service Registry for the Smart Living Ecosystem. In: INFORMATIK 2021. Gesellschaft für Informatik, pp. 1425-1438, 2021.

[Sc20] Schoormann, T.; Hagen, S.; Brinker, J.; Wildau, S.; Thomas, O.; Knackstedt, R.: Towards Aligning Business Models with Business Processes: A Tool-based Approach. In (Bork, D.; Karagiannis, D.; Mayr, H. C., eds.): Modellierung 2020. Gesellschaft für Informatik e.V., pp. 13–27, 2020.

[SF20] Sovacool, B. K.; Furszyfer Del Rio, D. D.: Smart home technologies in Europe: A critical review of concepts, benefits, risks and policies. In: Renewable and Sustainable Energy Reviews vol. 120, 2020.

[Va15] Valencia, A.; Mugge, R.; Schoormans, J. P. L.; Schifferstein, H. N. J.: The design of smart product-service systems (PSSs): An exploration of design characteristics. In: International Journal of Design vol. 9. IJDesign, pp. 13–28, 2015.

[WF94] Wasserman, S.; Faust, K.: Social network analysis: Methods and applications. Cambridge University Press, New York, 1994.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1681

Intelligente Hochschule – Digitalisierung und Künstliche Intelligenz im Hochschulkontext

- **Ausgewählte Aspekte -**

Ulrich John [1]

Abstract: Ohne Anspruch auf Vollständigkeit wird in Analogie zu Intelligenten Unternehmen motiviert, KI-Ansätze für das Management und die kontinuierliche Optimierung und Weiterentwicklung von Hochschulen zu entwickeln und zu verwenden. Durch die geeignete Integration von KI-Systemen/ KI-basierten Softwarekomponenten in adäquate Prozesse und zusätzliche Funktionalitäts- und Serviceerweiterungen ließen sich Agilität, Leistungsstärke, Qualität und Effizienz erheblich erhöhen. Einige diesbezügliche Aspekte, die ursprünglich als Bestandteile eines (studentischen) Forschungsprogramms „Intelligente Hochschule" definiert wurden, werden genannt und zur weiterführenden und ergänzenden Diskussion gestellt.

Keywords: KI-basierte Hochschule/ Intelligente Hochschule, Digitale Hochschule

1 Einleitung

Die Digitalisierung von Hochschulen ist seit mehreren Jahren in vollem Gange. Insbesondere durch die pandemiebedingten Herausforderungen an die Durchführung von Lehrveranstaltungen, Prüfungen, Besprechungen etc. haben sich digitale Plattformen für die Durchführung von online-Lehrveranstaltungen und die Präsentation und Verwaltung von Lehr- und Lernmaterialen unumkehrbar etabliert und bewährt. Dies trifft auch auf viele Fächer zu, bei denen die Durchführung von online-Lehrveranstaltungen vor der Pandemie eher auf Skepsis traf bzw. bei denen die Durchführbarkeit bezweifelt wurde (vgl. [FNS21]). Ebenso wie bei Unternehmen war diese IT-Unterstützung menschlicher Kommunikation nur durch die gereiften, einsatzfähigen IT-Technologien möglich. Hervorzuheben sind diesbezüglich insbesondere das Internet, Webtechnologien und das Cloud Computing. Viele bisherige Arbeiten konzentrieren sich schwerpunktmäßig auf die digitale Unterstützung, Virtualisierung und die digitale Anreicherung der Lehre, wobei neben gewünschter Skalierbarkeit und Kosteneinsparungen auch qualitätssteigernde Aspekte eine Rolle spielen. Einen breiteren Diskursbereich, der ebenso die Bereiche Forschung und Hochschulverwaltung umfasst, stellt Scheer unter dem Begriff *Hochschule 4.0*, recht allgemein, in [Sc18] vor. Hierbei empfiehlt er für Verwaltungsprozesse den Einsatz von Standardsoftware. Genau dies ist an mehreren Hochschulen derzeit zu beobachten, was – zu mindestens mittelfristig – diskussionswürdig ist, insbe-

[1] IU Internationale Hochschule, Frankfurter Allee 73A, 10247 Berlin, ulrich.john@iu.org/ john@intelligente-digitalisierung.de

doi:10.18420/inf2022_144

sondere wenn es sich um kommerzielle Standardsoftware mit laufenden Lizenzkosten handelt[2], die die Anforderungen nur zum Teil erfüllt. In Analogie zu Digitalen Unternehmen (vgl. [Jo15], [Jo17]) ist es wesentlich, bei jeglicher Form der Digitalisierung die gewünschte Funktionalität zu realisieren und dabei nachhaltig eine möglichst hohe Effizienz zu erreichen. Darüber hinaus sind die Agilität und eine berechtigungskonforme Transparenz zu sichern. Hinzu kommt, dass für einige Aufgabenbereiche, z.B. im Kontext der Hochschulverwaltung und Hochschulweiterentwicklung, keine/ kaum leistungsstarke Standardsoftware verfügbar ist. Generell sind für diesbezügliche Entscheidungsfragen die „üblichen" Abwägungen zwischen der Verwendung von „funktionalitätseinschränkender" Standardsoftware und der (kooperativen) Entwicklung von Individualsoftware durchzuführen.

Eine Hochschule ist ein komplexes System, das unter gegebenen Rahmenbedingungen gut funktionieren und sich zielgerichtet weiterentwickeln soll. Neben den derzeitig öfters thematisierten Digitalisierungsaspekten (siehe z. B. [FNS21]), die den Weg zur *Digitalen Hochschule* beschreiben, gestatten der Einsatz und die Integration KI-basierter Softwaresysteme/ Softwarekomponenten Leistungs- und Qualitätssteigerungen, die über die reine Unterstützung menschlicher Experten hinausgehen. Es bietet sich in diesem Zusammenhang an, dann den Begriff *Intelligente Hochschule* zu verwenden. Auch nach einer „KI-geprägten Digitalisierung" wird das System Hochschule maßgeblich bestimmt durch die menschlichen Akteure (Mitarbeiter/ -innen), die internen Prozesse und die prozessintegrierten Softwarefunktionalitäten. Das System verbleibt in einem zyklischen Weiterentwicklungsgesamtprozess, der signifikant von menschlicher Kreativität gespeist und gesteuert wird. Hinsichtlich der Gesamtgestaltung gibt es – abhängig vom jeweiligen technologischen Entwicklungsstand - Spielräume, die sich darauf beziehen, welche Tätigkeitsbereiche/ Funktionalitäten durch Mitarbeiter/ -innen und welche durch (KI-basierte) Softwarekomponenten übernommen werden können[3]. Die hierbei relevanten gesellschaftlichen Aspekte und auch die unternehmensethischen Fragestellungen sind wichtig (vgl. z.B. auch [Ge19]), sollen aber in diesem Paper nicht weiter thematisiert werden. Der Autor hat 2020 an seiner „vorherigen Hochschule" das (studentische) Forschungsprogramm „Intelligente Hochschule" ([Hw21]) entwickelt und initiiert[4], welches er inhaltlich an seiner jetzigen Hochschule (für studentische Arbeiten und Forschungstätigkeiten) einbringt und weiterverfolgen möchte. Die ursprüngliche Beschreibung des Programms (vgl. [Hw21], [Jo21]): „Das Programm ist in das Forschungsfeld „Agilität, Effizienz und Transparenz für komplexe Prozesse in Unternehmen, Unternehmensnetzen und Organisationen – Digitalisierung & Digitale Transformation" eingebettet, wobei im besonderen Maße Potenziale der sich entwickelnden Technologien der Künstlichen Intelligenz beleuchtet, thematisiert und genutzt werden sollen. Das inhaltliche Spektrum des Programms erstreckt sich von der Analyse bestehender Prozesse und Strukturen der Hochschule (ggf. auch der gesamten Unter-

[2] Unbestritten ist es sinnvoll, verfügbare Standardsoftware „übergangsweise" zu verwenden und sicher auch generell für einige Aufgabenbereiche.

[3] Die kreative Generierung neuer Ideen gehört sicher nicht dazu.

[4] An der bisherigen Hochschule wird das Programm wegen Hochschulwechsels des Initiators nicht weiter betrieben.

nehmensgruppe) bis zur Untersuchung der Einsatzmöglichkeiten (zu entwickelnder) KI-basierter Softwarekomponenten, dem Entwurf KI-basierter Hochschulprozesse und ggf. der prototypischen Realisierung, wobei die Aspekte Agilität, Transparenz und Effizienz eine wesentliche Rolle spielen sollen. Konkrete Teilgebiete sind zum Beispiel effektives Wissensmanagement, lernende Organisation und IT-basierte Reorganisation, Optimierung und Selbstorganisation, wobei neben reinen Prozess- und Technologieaspekten auch die Einbettung in gesellschaftliche Gesamtentwicklungen und der Faktor „Mensch" adressiert werden."

Im folgenden Kapitel werden kurz einige ausgewählte Aspekte der Digitalisierung von Hochschulfunktionalitäten thematisiert, die teilweise durch geeignete KI-basierte Softwarefunktionalitäten realisiert bzw. unterstützt werden können. Die vorstellten Teilthemen stellen keine abgeschlossenen Arbeiten dar und erheben keinen Anspruch auf Vollständigkeit. Sie sollen einer weiterführenden Diskussion und späteren Ergänzung dienen, um dann ggf. Zielstellungen für weiterführende Aktivitäten/ Projekte zu sein. Anschließend wird ein kurzes Fazit gegeben.

2 Digitale Hochschule / Intelligente Hochschule – ausgewählte Aspekte von Digitalisierung und KI-Einsatz

Hochschulen sind ähnlich wie viele Unternehmen, Organisationen und Behörden geprägt durch komplexe Prozesse und teilweise auch durch komplexe Strukturen. Es gibt verschiedene Zielaspekte für die Weiterentwicklung von Hochschulen, die auch abhängig vom Typ der Hochschule sind. Der Bildungsauftrag, den Hochschulen für sich proklamieren, impliziert, dass Studierende mit aktuellem (gesellschaftlich relevantem) Wissen versorgt werden und dass der Wissensvermittlungsprozess effizient und nachhaltig erfolgt. Neben dem Wissenserwerb steht natürlich insbesondere auch die Herausbildung von gewünschten Fähigkeiten der Studierenden im Fokus der Lehre. Auch als ergänzendes und bereicherndes Fundament der Lehre wird eine hochschuleigene Forschung betrieben und gefördert, die zur Weiterentwicklung der jeweiligen Fachgebiete beitragen soll und möglicherweise Relevanz für Wirtschaft und Gesellschaft hat. Neben den Digitalisierungsaspekten, die sich direkt auf die Lehre beziehen, wie z.B. die Möglichkeit, online-Lehrveranstaltungen durchführen zu können, gibt es allgemeine Ziele der Digitalisierung, die auch für Unternehmen, Organisationen und Behörden relevant sind: Digitale Unterstützung und ggf. Flexibilisierung der Prozesse, Beseitigung von Medienbrüchen, Informationslücken und „Blindarbeiten", Erhöhung der Prozesseffizienz und der Robustheit, Sicherung der Prozessqualität. Insbesondere bietet es sich auch an, Systeme zu schaffen, die Mitarbeiter/ -innen von Routinetätigkeiten befreien, wobei eben auch die Prozessgeschwindigkeit drastisch erhöht werden kann. Die freigesetzten Kapazitäten können dann anderen Tätigkeiten zugutekommen, wobei idealer Weise auch die Zufriedenheit der „betroffenen" Mitarbeiter/ -innen gesichert oder gesteigert werden kann.

Neben diesen allgemeinen Zielstellungen gibt es weitere Zielstellungen, die (s.o.) abhängig vom Typ der Hochschule sind und sich auch aus dem Einfluss und den Interessen der jeweiligen Stakeholder ergeben. Sicherlich gibt es hier zum Beispiel einige deutliche Unterschiede zwischen staatlichen Hochschulen und den meisten privaten Hochschulen. Unterschiedlich ausgeprägt können hierbei zum Beispiel die Aspekte Wachstum, Internationalisierung und Globalisierung sein. Eine private Hochschule kann zum Beispiel das Ziel haben, international an verschiedenen Standorten vertreten zu sein, wohingegen dies bei staatlichen Hochschulen in den überwiegenden Fällen keine Rolle spielt. Für jeden Hochschultyp/ für jede konkrete Hochschule lohnt sich im Kontext der Digitalisierung eine Stakeholder-Analyse. Bei allen Hochschulen sind natürlich die Studierenden, die in der Lehre, Forschung und Verwaltung tätigen Mitarbeiter/ -innen und die zukünftigen oder bereits derzeitigen (Duales Studium) Arbeitgeber relevante Stakeholder. Bei privaten Hochschulen sind zusätzlich die Gesellschafter maßgebliche Stakeholder, bei staatlichen Hochschulen der Staat oder/ und das jeweilige Bundesland[5]. Weitere Stakeholder sind existent, so z.B. Mitwettbewerber, Akkreditierungsagenturen, Beratungsunternehmen und Dienstleister.

Beispiele für allgemeine Digitalisierungsziele – im Einklang mit entsprechender Prozessgestaltung:

- Erhöhung der Service- und Ergebnisqualität bei gleichbleibender oder geringerer personeller (und infrastruktureller) Kapazität,
- passgenaue Erweiterung des Portfolios und Wachstum (mehr Standorte, mehr Studierende, mehr Kooperationsunternehmen),
- höherer Forschungs-Output.

Digitalisierungsmaßnahmen/ -ziele können mehr oder weniger stark durch die einzelnen Stakeholder bedingt bzw. auf sie orientiert sein. Im Folgenden werden einige kurz charakterisiert, d.h., eine ausführliche Erörterung/ Spezifikation unterbleibt, da dies nicht intendierter Inhalt des Papers ist.

2.1 Allgemeine Prozessunterstützung und Prozessautomatisierung

Ausgangspunkt hierfür ist eine Analyse der derzeitigen Prozesse und ggf. das Entwerfen von Zielprozessen. Dabei soll identifiziert werden, welche Teilprozesse digitalisiert/ automatisiert werden können/ sollen. Maßgeblich sind hierbei auch Erfahrungen bzgl. bis dato durchzuführender manueller Routinetätigkeiten, entstehende Fehlersituationen und erkannte Prozessrisiken. Im Zuge des Requirements Engineerings sollte natürlich nicht ausschließlich auf die Erfahrungen und Meinungen der Prozessbeteiligten geachtet werden. Neben der Tatsache, dass des Öfteren „Betriebsblindheit“ vorliegt, ist oft

[5] Staat/ Staaten und Bundesland/ Bundesländer sind natürlich auch Stakeholder bei privaten Hochschulen, jedoch anders priorisiert.

schlicht unbekannt und nicht vorstellbar, was durch Software geleistet werden kann.

Ein Blick in die Praxis zeigt zum Beispiel: Überall, wo Formulare mit Informationen ausgefüllt werden müssen, die „an anderen Stellen vorliegen“ gibt es überwindbare Medienbrüche. Die Behebung dieser führt i.d.R. in Summe zu nicht zu unterschätzenden Zeiteinsparungen und zu höheren Zufriedenheitswerten bei den bisher betroffenen Mitarbeitern/ -innen. Hierbei handelt es sich im Wesentlichen um „reine Digitalisierungen“, die keine KI-basierten Komponenten benötigen/ verwenden. Zielführend ist in diesem Kontext ebenfalls eine Analyse bisheriger Informationsflüsse einschließlich identifizierter Mängel und Fehlersituationen[6]. Hierbei ist schlicht das informationslogistische Prinzip zu sichern bzw. umzusetzen. Relevante Informationen müssen die entscheidenden Akteure fehlerfrei zum richtigen Zeitpunkt erreichen. Wird zum Beispiel die Prüfungsart in einem Lehrmodul geändert, müssen die betroffenen Stakeholder nachhaltig informiert werden bzw. muss systemseitig gesichert werden, dass nicht Prüfungen auf Basis der bisherigen Prüfungsart vorbereitet werden können und so weiter. Bezüglich der zu realisierenden Informationsflüsse ist eine entsprechende Kommunikationsplanung durchzuführen, die unter anderem sichert, dass Stakeholder nicht mit überflüssigen Informationen „zugeschüttet“ werden, sondern dass nur die relevanten Informationen auf geeignete Art und Weise kommuniziert werden, z.B. Festlegung für die einzelnen Informationskategorien und Stakeholdergruppen oder gar individuelle Stakeholder, ob das push- oder das pull-Prinzip verwendet werden soll. Nicht alles, was kosten- und zeiteinsparend digitalisierbar ist, muss zwingenderweise digitalisiert werden! Diesbezüglich sind entsprechende Abwägungen durchzuführen.

In der Regel verwenden Hochschulen bereits verschiedenste Informations- und Verwaltungssysteme. Werden diese/ solche geändert oder neu realisiert, sind in jedem Fall die aus dem Software Engineering bekannten Prinzipien zu berücksichtigen, zum Beispiel ausreichendes *Testen* und *korrekte durchgehende Dokumentation*. Nur so können spätere Erweiterungen und Korrekturen aufwandsarm durchgeführt werden. Bei allen inhaltlichen durch die Software abzubildenden Regeln und Randbedingungen ist zu hinterfragen, ob diese dauerhaft gelten. Andernfalls ist eine Parametrierung bzw. deklarative Beschreibung und Adaptierbarkeit vorzusehen. Regelungen, z.B. bzgl. der Anzahl erlaubter Wiederholungsprüfungen während eines Studiums können sich ändern!

Grundlage für ein schlüssiges Informationsmanagement-System einer Hochschule wird i.d.R. *eine* Datenbank sein sollen, die dann natürlich entsprechend der zu erfüllenden Anforderungen konzipiert sein muss. Erfahrungsgemäß werden an einigen Hochschulen immer noch mehrere Datenbanken mit sich überschneidenden Informationen verwendet, was die Gefahr von Medienbrüchen und ungewünschten Inkonsistenzen birgt. Auf Details soll an dieser Stelle nicht eingegangen werden.

[6] Wenn nicht bereits geschehen, sollte die Hochschule zu einer lernenden Organisation gewandelt werden. Das heißt insbesondere: Fehler und erkannte Verbesserungspotenziale werden geeignet erfasst, aufbereitet und für Prozess- sowie ggf. Strukturverbesserungen genutzt.

2.2 Bill of Study & erweiterte Leistungsbilanz

Mit Durchführung/ Umsetzung der Bologna-Reform ergaben sich für (zukünftige) Arbeitgeber der Studierenden erhebliche Unsicherheiten hinsichtlich der Bewertbarkeit und Belastbarkeit von Abschlussnoten absolvierter Kurse. Dies resultiert aus der Tatsache, dass Prüfungsleistungen, im Vergleich zu früher überwiegend mündlichen durchgeführten Prüfungen und Klausuren, aufgeweicht wurden, um die Belastung der Studierenden zu reduzieren. Für einen Arbeitgeber ist es unter Umständen extrem relevant, zu wissen, welcher Kurs durch welche Art von Prüfungsleistung benotet wurde. Eine für ein Fachgebiet auf Grundlage eines Referats oder einer Posterpräsentation vergebene Note sagt nichts oder nur sehr wenig über die tatsächlichen Kenntnisse des/ der Studierenden auf diesem Fachgebiet aus. Gerade in Zeiten dringender Digitalisierungen in Wirtschaft und Gesellschaft birgt diese Tatsache erhebliche Risiken für den jeweiligen Erfolg. Die Idee ist, die entsprechenden Informationen (optional und für die Studierenden freiwillig) auf einer *Bill of Study* auszuweisen[7]. Zusätzlich zu der jeweiligen Prüfungsart könnte dort beispielsweise vermerkt sein, welche thematischen Inhalte der Kurs hatte, bei welchem/ welcher Dozenten/ -in die/ der Studierende welchen Kurs absolviert hat, welche Durchschnittsnote dieser Kurs hatte und wie oft er/ sie an den jeweiligen Kurslehrveranstaltungen teilgenommen hat. Zusätzlich könnten zusammenfassende Informationen über alle absolvierten Lehrveranstaltungen softwaregestützt auswertbar angegeben sein, wie z.B. die *tatsächlich* während des gesamten Studiums behandelten Themen, ggf. sogar bezogen auf die vom Studierenden konkret besuchten Lehrveranstaltungen. Dazu müssten von der Hochschule entsprechende Informationen (automatisch) erfasst und verwaltet werden. Natürlich sind hierfür datenschutzrelevante Aspekte zu berücksichtigen[8]. In jedem Fall wäre ein solcher zusätzlicher Service eine für die Stakeholdergruppe *Arbeitgeber*, für den Staat und ggf. auch für die Studierenden nutzbringend. Werden solche erweiterten Leistungsbilanzen elektronisch auswertbar erstellt, könnten diese auch von fachfremden Arbeitgebern über Inanspruchnahme entsprechender Dienstleistungen bewertet und eingeschätzt werden.

2.3 Serviceportale für Stakeholder & Informationsaufbereitung

Wie bereits für das Thema „Bill of Study“ als Möglichkeit genannt, gäbe es auch Möglichkeiten, für andere Stakeholder digitale Services zu schaffen und zu nutzen oder sogar Informationsaustausch über zu realisierende Systemschnittstellen zu ermöglichen. Auf Portale für Studierende und Mitarbeiter/ -innen der Hochschule, die heutzutage bereits verbreitet im Einsatz sind, soll hier nicht weiter eingegangen werden. Ein Beispiel für externe Stakeholder ist das Statistische Bundesamt, dem alle deutschen Hochschulen regelmäßig detaillierte Informationen/ statistische Daten zuliefern müssen. Statt hierfür Mitarbeiter/ -innen einzusetzen und entsprechende zyklische Kommunikationsprozesse

[7] Alternativ könnte dies nach Freigabe des jeweiligen Studierenden für “berechtigte Interessenten“ von der Hochschule als Service angeboten werden.

[8] Optional könnte den Studierenden ermöglicht werden, zusätzliche Themen einzuspeisen, die sie sich im Selbststudium für das Fach erschlossen haben.

innerhalb der Hochschule aufzusetzen, sollten die Informationen systemseitig aktuell gehalten und automatisch aufbereitet werden, so dass eine zyklische Datenlieferung an das Statistische Bundesamt automatisch erfolgen kann.

Mit der Zielstellung, Win-Win-Effekte zu erzielen, könnten ebenfalls aufbereitete Informationen (digital verarbeitbar) darüber angeboten werden, welche Dozierenden in welchen Semestern welche Kurse in welchen Studiengängen mit welchen Kursinhalten durchführen. Dies könnte beispielsweise für kooperierende Hochschulen, Akkreditierungsagenturen, Projektträger aber vor allem auch für Fachbuchverlage und potenzielle (industrielle) Forschungsauftraggeber interessant sein. Die Fachbuchverlage könnten gezielt Neuerscheinungen „platzieren“ und auf der anderen Seite würden die Dozierenden Zeit sparen können und mit aktuellen Informationen versorgt werden.

Sicher sind noch weitere sinnvolle Services denkbar.

2.4 Automatische Generierung und Aktualisierung von Lehrmaterialien

Ergänzende und unterstützende Lehrmaterialien besitzen, in konventioneller bewährter Form als Skripte und Präsentationen aber auch in modernerer Form als systemseitig auswertbare Übungsaufgaben, Webkurse etc., große Bedeutung. Wesentlich hierbei sind die Qualität und Aktualität. Generell zu durchdenken wäre im Kontext der Hochschuldigitalisierung, inwieweit zumindest Teile dieser Lehrmaterialien automatisch oder mindestens halbautomatisch aus verfügbaren Quellen (z.B. mittels unterstützender Web-Dämonen) generiert und aktualisiert werden können. Das macht insbesondere darum Sinn, weil es sich bei der rein manuellen Erstellung und Aktualisierung um einen zeitaufwendigen Prozess handelt.

Eine automatische Aktualisierung von Daten (z.B. aktuelle Börsenwerte und andere aktuelle Kennzahlen) in Präsentationen und Skripten ist realisierbar und wird – einschließlich der automatisierten Generierung verschiedenster Diagrammtypen - bereits sehr erfolgreich kommerziell vermarktet[9]. Weiterhin könnten zumindest automatisiert Änderungsmeldungen an die Autoren/ -innen der Lehrmaterialien geschickt werden, wenn bei angegebenen themenbezogenen Webseiten Änderungen durchgeführt wurden[10], wobei die entsprechenden Änderungen berechnet und geeignet visualisiert werden können. Höherwertige Aufbereitungen, Extraktionen und Zusammenfassungen[11] von vorliegenden Informationen aus verschiedenen Quellen sind perspektivisch ein Einsatzgebiet für KI-basierte Verfahren. Anregungen hierfür können aus anderen Gebieten, z.B. der Content-Erkennung in juristischen Dokumenten, gewonnen werden. Für automatisiertes Layouten könnten darüber hinaus constraint-basierte Ansätze verwendet werden.

[9] siehe z.B. www.think-cell.com
[10] Analoges gilt für forschungsrelevante Informationen.
[11] Hier könnten ggf. auch geeignete Serviceprovider eingebunden werden.

2.5 Wissensmanagement

Digital unterstütztes Wissensmanagement für verschiedenste Stakeholdergruppen und Themenbereiche ist ein wichtiges zukunftsträchtiges Thema für die Weiterentwicklung von Hochschulen, das sicherlich weit über Dokumentenmanagement, Mindmaps und Wikis hinausgehen kann. Neben der Schaffung geeigneter Softwaresysteme, die ggf. auch Data & Knowledge Mining für verschiedene Teilbereiche unterstützen könnten, sind natürlich die Schaffung geeigneter/ kompatibler Prozesse (z. B. zum Lernen aus Fehlern und Problemen & zur Unterstützung der Phasen des SECI-Modells) sowie die Integration von Social-Media-Komponenten von entscheidender Bedeutung. Diesbezügliche Anregungen bietet unter anderem [Bä16].

Die Wissensmanagementkomponenten sollen allen Stakeholdergruppen einen raschen Zugriff auf relevantes, korrektes Wissen erlauben und darüber hinaus die Möglichkeit bieten, Anreicherungen der Wissensbasen oder gar eigene oder kooperative Wissensbasen zu erstellen und weiterzuentwickeln. Hierfür sind verschiedenste Bausteine denkbar. Ein eher trivialer ist die Erstellung und Pflege von FAQs und von geeigneten Prozessdokumentationen. Unterstützt durch den Einsatz symbolischer KI-Systeme könnten aus individuellen und kooperativen Wissensbasen Ontologien und semantische Netze erstellt werden.

Perspektivisch ist sicherlich die Entwicklung und der Einsatz von Sprachinterfaces[12] sinnvoll, sofern Sie eine über Trivialanfragen hinausgehende Funktionalität anbieten.

Sicht der Stakeholdergruppen

Neben den kurz geschilderten Funktionalitätsanforderungen wäre aus Sicht der Studierenden sicher der zu Ihrem Studiengang/ Ihren Studiengängen passende Aufbau einer Wissensbasis wünschenswert. Einerseits sollten Funktionalitäten bereitgestellt werden, die es den Studierenden ermöglichen, sich in Bezug zu ihren Lehrveranstaltungen, eigene oder kooperative Wissensbasen zu erstellen. Andererseits könnten solche Wissensbasen – ggf. unter sukzessiver Verwendung aufbereiteter studentischer Wissensbasen – fachbezogen im Laufe der Zeit erstellt und ergänzend zu anderen Lehrmaterialien zur Verfügung gestellt werden.

Ähnliches gilt für Lehrende, fachbezogen und lehrbezogen, bezogen auf Forschungsthemen/ -felder, bezogen auf die akademische Selbstverwaltung sowie für nichtprofessorale Mitarbeiter/ -innen der verschiedensten Hochschulbereiche.

Ein Beispiel für eine sehr nutzbringende Funktionalität für Lehrende und Studierende wäre in diesem Zusammenhang die Möglichkeit, tatsächliche Inhalte (und Detailierungsgrade) durchgeführter Lehrveranstaltungen erfassbar zu machen. Eine systemunterstütze Auswertbarkeit ist insbesondere für thematisch verbundene Kurse und Folgekurse von Interesse. Durch entsprechende Berücksichtigung kann durch alle beteiligten

[12] Sollen hierfür Web-Services externer Anbieter verwendet werden, müssen Unbedenklichkeits-Untersuchungen bzgl. der Informationssicherheit und des Datenschutzes durchgeführt werden.

Lehrenden besser gesichert werden, dass die Studierenden eines Studienganges zum Ende ihres Studiums tatsächlich alle vorgesehenen Themenkomplexe mit adäquatem Detailierungsgrad behandelt haben und das darüber hinaus ungewollte thematische Wiederholungen vermieden werden. Von Interesse ist dies natürlich auch für die effektive Planung und Durchführung konsekutiver Studiengänge. In diesem Zusammenhang wäre perspektivisch auch eine Konfigurationsfunktionalität für die Unterstützung bei der Inhalts- und Durchführungsplanung von Lehrveranstaltungen denkbar.

2.6 Anlauf- und Auslaufplanung von Studiengängen

Insbesondere bei großen Hochschulen sind beträchtliche Planungsprobleme zu lösen, die sich relativ gut mit CP-basierten/ CLP-basierten[13] Modellen repräsentieren lassen. Das bedeutet jedoch nicht per se, dass für die komplexen Planungsprobleme auch in kurzer Entwicklungszeit praktikable Planungskomponenten entwickelt werden können. Vielmehr ist zu erwarten, dass – zumindest für einige Planungsproblemtypen ab einer bestimmten Problemgröße - entsprechend Zeit und Know-How in die Modellierung und Entwicklung von Suchheuristiken sowie in das Finden und Testen redundanter suchraumeinschränkender Constraints zu investieren sind[14]. Sind diese Arbeiten von Erfolg gekrönt, wird man systemunterstützt deutlich bessere Pläne in deutlich kürzerer Zeit berechnen können, als dies durch menschliche Planer möglich wäre. Die entsprechenden Softwarekomponenten können bedingt durch ihre inhärente „CP-Natur“ hochflexible Interaktionen mit den menschlichen Planern/ Planerinnen ermöglichen, sodass einerseits nicht formalisiertes Expertenwissen in die Planung einfließen kann und andererseits verschiedene What-If-Analysen durchgeführt werden können.

Eine hochschulrelevante Planungsproblemklasse, die insbesondere auch mit der Lehrkapazitäts- und Aufwuchsplanung verbunden ist, ist der Anlauf und Auslauf von Studiengängen. Abhängig von den erlaubten/ sinnvollen Gruppengrößen und den Möglichkeiten, aufgrund der online-Durchführbarkeit einzelner Lehrveranstaltungen, Studierendengruppen verschiedener Standorte[15] für einige Lehrveranstaltungen zusammen zu legen, ist abhängig von verfügbaren Lehrkapazitäten zu klären, ob Module oder Teilmodule verschiedener Studiengänge aufgrund inhaltlicher Identität (vorübergehend) zusammen gelehrt werden können.

2.7 Konzeption/ Design von Studiengängen

Wesentlich für den Erfolg einer Hochschule ist das Angebot inhaltlich attraktiver

[13] CP = Constraint Programming, CLP = Constraint-Logic Programming.

[14] Möglicherweise ist es für einige Problemgrößen notwendig, hybride Ansätze mit anderen Planungs- und Optimierungstechnologien zu entwickeln, z.B. unter Einbeziehung von evolutionären oder genetischen Algorithmen.

[15] Bei vorliegenden Gegebenheiten macht dies bereits auch für die Planung verschiedener Studiengänge an einem Standort Sinn.

Studiengänge. Die Entwicklung eines neuen Studiengangs erfordert i.d.R. einen erheblichen Aufwand. Analog zu dem bereits im vergangenen Abschnitt motivierten Vorgehen, inhaltlich passende Module möglichst in mehreren Studiengängen zu nutzen, entstehen hierbei Planungs- und Konfigurations-/ Designprobleme, die darin bestehen, möglichst viele attraktive Studiengänge mit einer möglichst großen Anzahl „gemeinsam genutzter" Module zu konzipieren, wobei ggf. auch (perspektivische) Kapazitäts- und Ressourcenverfügbarkeits-Constraints berücksichtigt werden sollen. Weiterführende Informationen und Ideen zu diesem Thema findet man in [Jo14].

2.8 Erweiterte Lehrveranstaltungsplanung/ Timetabling

Die Lehrveranstaltungsplanung (Timetabling) für Hochschulen ist eine hochkomplexe Planungsproblemklasse, die ebenfalls mit CP/ CLP-basierten Softwaremodellen adressiert und realisiert werden kann (z.B. [Go06]). Neben der Verfügbarkeit von Ressourcen (Lehrende, Räume entsprechender Größe, Studierende[16], Speziallabore) sind weitere Randbedingungen zu berücksichtigen, zum Beispiel zeitliche Reihenfolgebeziehungen zwischen Lehrveranstaltungen aufgrund inhaltlicher Verzahnungen der zu planenden Kurse, die Erfüllung des Quorums, Eignungen von Fächern für die online-Lehre und ggf. gewünschte Verhältnisse zwischen online- und Präsenzlehre in den verschiedenen Studiengängen und Studiensemestern. Mit der Möglichkeit der online-Durchführbarkeit von Lehrveranstaltungen ergibt sich eine weitere Planungsdimension, da passend zu den Randbedingungen (s.o.) standortübergreifend[17] für einige Kurse „gemeinsame Lehrveranstaltungen durchgeführt werden können. Dies macht insbesondere bei kleineren Gruppen oder bei Lehrkraftengpässen Sinn oder auch bei bestehendem Wunsch, die Studierenden gezielt standortübergreifend zu verlinken, was insbesondere bei Dualstudierenden eine Wissens- und Erfahrungsverbreiterung begünstigt.

2.9 Unterstützung für Akkreditierungen

Akkreditierungen und Re-Akkreditierungen sind sehr zeit- und ressourcenaufwendig. Neben einer weitergehenden Qualitätssicherung könnten die dafür erforderlichen Aufwände durch entsprechende Digitalisierung erheblich verringert werden. Würden zum Beispiel die Modulkataloge der verschiedenen Studiengänge formalisiert[18] werden (das bräuchte man auch für das Design neuer Studiengänge, vgl. Abschnitt 2.7), könnten automatische Konsistenzchecks bzgl. der logischen Verzahnung einzelner Module durchgeführt werden. Außerdem könnten über etwaige Schnittstellen automatisch neue Auflagen von Referenzliteratur in die Modulkataloge und annotierte Dokumente/ Webseiten etc. „eingepflegt werden", ggf. unter zusätzlicher automatischer Benachrichtigung der jeweiligen Modulverantwortlichen und der entsprechenden Dokumentation.

[16] In Dualen Studiengängen sind die Studierenden nicht an allen Wochentagen verfügbar.
[17] Falls die Hochschule über mehrere Standorte verfügt.
[18] semantisch auswertbare Repräsentation

Neben der automatischen Aktualisierung der Literaturdaten[19], könnten verbundene Verlage (vgl. auch Abschnitt 2.3) oder auch entsprechend parametrierte Web-Crawler gezielt Neuerscheinungen zu den Themengebieten der einzelnen Module vorschlagen, die dann bei Bedarf durch die jeweiligen Modulbeauftragten hinsichtlich der Aufnahme in die Literaturempfehlungen begutachtet werden können.

Entsprechende Web-Crawler könnten in diesem Kontext auch konzipiert und realisiert werden, um etwa - themengebietsbezogen - verschiedene Webseiten, z.B. von anderen Hochschulen, Unternehmen, Konferenzen und Behörden zu „überwachen“ und Änderungen dann automatisch aufzubereiten und zu kommunizieren.

Somit ließe sich insgesamt aus Sicht der Hochschulen eine systemseitige Unterstützung konzipieren und realisieren, die die Qualität der Modulkataloge sichert und den erforderlichen Aufwand deutlich reduziert.

Aus Sicht des Staates und möglicherweise auch aus Sicht der Akkreditierungsagenturen wären formalisierte Studiengangbeschreibungen ebenfalls wünschenswert, da entsprechende Prüfungen bei der Durchführung von Akkreditierungen dann ebenfalls – zumindest teilweise – automatisiert durchgeführt und dokumentiert werden könnten. Bei diesen automatischen/ teilautomatischen Begutachtungen könnten Studiengangs- und fachspezifische Informationen mit Informationen von anderen Akkreditierungen (anderer Hochschulen) abgeglichen werden. Es könnte also hier ein Lernen realisiert werden, welches zur Generierung von Verbesserungsvorschlägen und damit zur Erhöhung der Prozess- und Inhaltsqualität genutzt werden kann.

2.10 Outsourcing/ Schaffung von Serviceprovidern & Kooperationen

Aus Sicht des Staates bzw. der Länder und ggf. auch aus Sicht einiger privater Hochschulen, könnte es sinnvoll sein, zu überlegen, ob einzelne Prozesse oder Funktionsbereiche nach einer Digitalisierung gebündelt oder/ und als Service angeboten werden können. Ist dies ebenfalls ein interessanter Gedanke für Start-ups? So könnte zum Beispiel ein zertifizierter Serviceprovider (ggf. unter Nutzung von GAIA-X-Funktionalitäten) in der Cloud virtuelle, parametrierbare Prüfungsämter anbieten.

Ein anderer Aspekt sind systemunterstützte, erweiterte Kooperationen zwischen Hochschulen. Durch Schaffung entsprechender Systeme, z.B. geeignet konzipierte kommunizierende Systeme oder entsprechende Cloud-Plattformen ließen sich einerseits Kollaborations-/ Kooperationsprozesse unterstützen, andererseits ließen sich neue Dimensionen von Kooperationen realisieren. In diesem Zusammenhang könnte es zum Beispiel interessant sein, über die gemeinsame Planung und Durchführung oder auch über die Auftragsdurchführung von online-Kursen oder ganzen Studiengängen nachzudenken,

[19] Sinnvoll wäre hier auch eine automatische Prüfung, ob relevante Themenbereiche in den neuen Auflagen entfallen sind. Hierzu müssten zum Beispiel entsprechende, automatische Kommunikationen mit Fachbuchverlagen realisiert werden.

was natürlich dann Auswirkungen auf die planerischen Randbedingungen der beteiligten Hochschulen hat, was wiederum zu systemunter-stützten, kollaborativen Planungsprozessen führen könnte.

3 Fazit

Basierend auf dem 2020 vom Autor initiierten Forschungsprogramm „Intelligente Hochschule“ wurden – ohne Anspruch auf Vollständigkeit und adäquate Gewichtung – mögliche Einsatzbereiche und Aspekte für weiterführende (KI-basierte) Digitalisierung und Digitale Weiterentwicklung von Hochschulen benannt und andiskutiert. Absicht ist, einzelne Teilthemen, z.B. in Form von studentischen Projekten und/ oder Abschlussarbeiten zu untersuchen und weiterzuentwickeln, ggf. in Kooperation mit anderen Hochschulen/ Universitäten oder/ und anderen Stakeholdergruppen.

Verbunden mit den diskutierten Aspekten ist das Thema *Digitaler Hochschulzwilling*. Sicher ist ein solcher - in feingranularer Form – in absehbarer Zeit nicht realisierbar. Lohnenswert wäre aber auch die Schaffung eines Zwillings (für ausgewählte Bereiche) auf einem praktikablen Abstraktionsniveau, sodass für wesentliche Fragestellungen des Hochschulmanagements und der Hochschulentwicklung ein mulitdirektional auswertbares Modell – im Sinne einer Entscheidungsunterstützung - zur Verfügung stehen würde, das darüber hinaus unter Verwendung von im Laufe der Zeit generierten Daten und gesammelten Erfahrungen entsprechend unterfüttert und weiterentwickelt werden könnte. Neben der Planungsunterstützung durch ein solches Modell könnte der Hochschulzwilling ebenfalls Grundlage für die annotierte Dokumentation des Hochschulbetriebs und der Hochschulentwicklung sein.

Literaturverzeichnis

[Bä16] Bächle, M.: Wissensmanagement mit Social Media – Grundlagen und Anwendungen. De Gruyter Oldenbourg, 2016.

[FNS21] Friedrich, J.-D., Neubert, P., Sames, J.: 9 Mythen des digitalen Wandels in der Hochschulbildung. Diskussionspapier Nr. 13. Berlin: Hochschulforum Digitalisierung.

[Ge19] Geske, U.: Künstliche Intelligenz und Digitalisierung – sind wir auf dem richtigen Weg? In [JHW19].

[Go06] H.-J. Goltz: Course Timetabling and Automatic Assignments of Students to Courses. Proc. CHIP User Club Meeting, Paris, April 2006.

[Hw21] hwtk: HWTK Forschung im Fokus, Forschungsbericht Akademisches Jahr 2019/ 20, Berlin, 2021.

[JHW19] John, U., Hofstedt, P., Wolf, A. (Hrsg.): Deklarative Ansätze zur Künstlichen Intelligenz – punktuelle Beiträge. Online-Post-Proceedings des 32nd Workshop on (Constraint) Logic Programming, Berlin, 2019.

[Jo14] John, U.: Design and Redesign of University Course Programs. 28. Workshop „Planen, Scheduling und Konfigurieren, Entwerfen“ (PuK 2014) auf der KI 2014. Stuttgart, 2014.

[Jo15] John, U: Digitales Unternehmen – Bausteine für Effizienz, Agilität und Transparenz. INFORMATIK 2015 (Cunningham, D. et al (Ed.)), Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, 2015.

[Jo17] John, U.: Deklarative Unternehmensmodelle – essentielle Bausteine Digitaler Unternehmen für optimierende Reorganisationen. INFORMATIK 2017 (Maximilian Eibl, Martin Gaedke (Ed.)), Lecture Notes in Informatics (LNI), Gesellschaft für Informatik 2017.

[Jo21] John, U.: Ein Forschungsfeld stellt sich vor: Agilität, Effizienz und Transparenz für komplexe Prozesse in Unternehmen, Unternehmensnetzen und Organisationen – Digitalisierung und Digitale Transformation. In [Hw21].

[Sc18] Scheer, A.-W.: Unternehmung 4.0 – Vom disruptiven Geschäftsmodell zur Automatisierung der Geschäftsprozesse. AWSi Publishing, 2018.

D. Demmler, D. Krupka, H. Federrath. (Hrsg.): INFORMATIK 2022,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2022 1695

Intelligentes Unternehmen, Intelligenter Staat, Intelligentes Europa – Was ist zu tun?

- Anstoß einer kollegialen Experten-Diskussion -

Ulrich John [1]

Abstract: Ziel des bereits 2021 geplanten Kurzpapers ist der Anstoß einer Diskussion unter Fachkollegen/ -innen über die aktuelle Situation und über Möglichkeiten, die dringend notwendigen Digitalisierungen zu sichern bzw. „auf den richtigen Weg zu bringen". Geplant ist, eine diesbezügliche Initialdiskussion auf dem 14. Workshop {KI-basiertes} Management und Optimierung komplexer Systeme (MOC 2022) zu führen. Ergebnisse wären idealer Weise der Anstoß von Folgediskussionen und zielführenden Aktivitäten bzw. Initiativen.

Keywords: Intelligentes Unternehmen, Intelligenter Staat, Intelligentes Europa, Intelligente Digitalisierung von Unternehmen, Organisationen, Behörden und Verbünden

1 Aspekte der derzeitigen Situation

Digitalisierung und Digitale Transformation sind weltweit nach wie vor zentrale, beherrschende Themen- und Herausforderungsgebiete in Wirtschaft, Politik und Gesellschaft. Unternehmen, Staaten und Staatengemeinschaften, denen es gelingt, ihre Prozesse und Strukturen adäquat an den jeweiligen Zielen auszurichten, diese effizient zu gestalten sowie zielgruppenkonforme Transparenz und Agilität zu sichern, werden das „Rennen machen". Die Möglichkeit, Unternehmen, Behörden und Verbünde weitgehend zu digitalisieren, basiert auf den gereiften IT-Technologien der letzten Jahrzehnte. Neben der IT-basierten Automatisierung der Prozesse (vgl. z.B. [Sc18]) bieten weiterentwickelte und gereifte KI-Technologien in Kombination mit anderen Technologien, wie z.B. dem Cloud-Computing, die Grundlage für eine weitere, aktuelle Welle der Digitalisierung, die enorme Potenziale für Servicegrad- und Performanzsteigerungen besitzt sowie Kosten drastisch reduzieren kann. In zunehmendem Maße begreifen Unternehmen und Staaten, dass insbesondere das Gebiet der Künstlichen Intelligenz entscheidend für den Erhalt der Position oder die (Neu-)Positionierung auf dem Markt und im Wettbewerb der Staaten und politischen Systeme ist (vgl. z.B. [Si19]). Spät wird die in relativ kurzer Zeit aufgebaute und sich rasant weiter entwickelnde „Macht" der großen Internetkonzerne erkannt. Die technologischen, wirtschaftlichen und politischen Entwicklungen sowie der weltweite Macht- und Einflussausbau Chinas und die Verletzlichkeit der bestehenden Handels- und Sicherheitsbeziehungen werden mit großer Sorge

[1]IU Internationale Hochschule, Frankfurter Allee 73A, 10247 Berlin, ulrich.john@iu.org/ john@intelligente-digitalisierung.de

doi:10.18420/inf2022_145

betrachtet. Oft liegt der thematisierte Hauptfokus auch auf dem militärischen und machtpolitischen Bereich (Cyberwar, Wahlbeeinflussungen etc.). Unbenommen der Relevanz dieses Bereiches sollte der geeignete Einsatz von KI für alle Bereiche von Wirtschaft und Gesellschaft untersucht und thematisiert werden. Egal, welche Ebene (Unternehmen, Landesbehörde, Bundesbehörde, Staat, Staatengemeinschaft, G7 etc.) man betrachtet: Es gilt, die Herausforderungen, Chancen und Risiken zu erkennen und möglichst unverzögert und transparent „die richtigen Dinge" zu tun. Es scheint die Gefahr zu bestehen, dass sich Teile von Wirtschaft, Politik und Gesellschaft derzeit mit einzelnen Teilaspekten der (KI-basierten) Digitalisierung verzetteln[2]. In diesem Paper soll bewusst nicht konkret auf ausgewählte Digitalisierungs- und KI-Initiativen, Programme, Projekte und Gremien in Deutschland und der EU, wie Digitalkabinett, Digitalrat, IT-Rat, IT-Planungsrat, KI-Strategie der Bundesregierung, Europäisches KI-Gesetzt etc., eingegangen werden, obwohl eine genaue Analyse in einem anderen Kontext notwendig und erkenntnisreich wäre. Für einige diesbezügliche, insbesondere KI betreffende, Überlegungen/ Aspekte sei auch auf [Ge19] verwiesen.

Leider hat sich in den ersten Monaten der Pandemie gezeigt, dass viele „Dinge" in Deutschland und der Europäischen Union nicht befriedigend funktioniert haben. Unverständlich ist zum Beispiel, warum ein als technologisch hochentwickelt geltendes Land, wie Deutschland, nicht ausreichend vorbereitet war, relativ „gering-komplexe" Produkte, wie Atemschutzmasken in kurzer Zeit in großer Anzahl zu produzieren, warum wurde oft viel zu spät zielführend reagiert. Ebenfalls sichtbar wurden nun auch Folgen der in den letzten Jahren durchgeführten marktgetriebenen Auslagerung relevanter Fertigungsstufen verschiedenster wichtiger Produkte (z.B. Medikamente) im Kontext der Globalisierung. Trotz jahrelanger Warnungen vor globalen Pandemien hat hier das System „Staat" nicht ausreichend funktioniert. (Digital unterstützte) Prozess- und Ressourcenabsicherungen im Rahmen eines Risikomanagements wären hier offensichtlich „angebracht gewesen"[3]. In diesem Zusammenhang stellen sich (erneut) einige weitere Fragen. Zum Beispiel ist zu klären, wieviel Verantwortung der Staat übernehmen soll und muss. Ist es wirklich legitim, essentielle Leistungserbringungen und Basisdienste[4] allein den am Markt agierenden Unternehmen zu überlassen und sich auf das Funktionieren zu verlassen? Muss die Gestaltung von Lieferketten für systemrelevante Produkte – abgesehen von den Regelungen des Lieferkettentransparenzgesetzes – vom Gesetzgeber reglementiert werden? Diese Frage stell sich nun erneut im Kontext des Ukrainekriegs. Ist systemisch gesichert, dass die erforderlichen Digitalisierungsmaßnahmen/ -projekte zielgerecht und erfolgreich durchgeführt werden können? Existiert eine ausreichende berechtigungskonforme Transparenz? Diese Fragen lassen sich analog auf Unternehmen, die EU usw. übertragen/ anpassen.

[2] Interessante Aspekte bzgl. der Bewertungen der derzeitigen Situation findet man z.B. auch in [Me21].

[3] Fairer Weise muss erwähnt werden, dass einige Dinge (mit zeitlicher Verzögerung) unter teilweise erheblichen persönlichen Einsatz einiger Akteure relativ gut funktionieren.

[4] verbunden mit regelmäßigen enormen Lizenzgebühren z.B. an Microsoft, allein für das Jahr 2020 hat der Bund ca. 178 Mio. Euro für Lizenzen an Microsoft gezahlt (vgl. z.B. Tagesspiegel 18.5.21), 2021 ca. 205 Mio. EUR (vgl. heise.de 26.2.22)

Neben den zu Tage tretenden Defiziten wird bescheinigt, dass die bisherigen[5] Digitalisierungsaktivitäten auf Landes- und Bundesebene „stark verbesserungswürdig“ und „eher unbefriedigend“ sind. Der Rechnungshof des Landes Berlin bescheinigte 2021 ([Re21]), dass trotz mehrerer Jahre Projektarbeit auf dem Gebiet der „Landes-Digitalisierung“ eine *Intransparenz über die IT-Systeme* herrscht, die in den 9 A-Landesbehörden und den 58 B- und C-Landesbehörden im Einsatz sind. Ähnliches lässt sich für die damit verbundenen Prozesse vermuten. Angemahnt wird auch eine unzureichende Personalbedarfsermittlung in Behörden des Landes Berlin. Ähnlich gelagerte Kritiken werden auch auf Bundesebene, z.B. im Gutachten „Digitalisierung in Deutschland – Lehren aus der Corona-Krise“ ([BM21]) des Wissenschaftlichen Beirats beim Bundesministerium für Wirtschaft und Energie, formuliert. Dort kann man unter anderem lesen, dass Deutschland beim E-Government innerhalb der EU (27 Mitgliedsstaaten) auf Platz 21 rangiert! Das kann Angst machen und wir müssen zügig die Ursachen erkennen und analysieren, um konstruktiv gegenzusteuern. Die Frage nach systemischen Gründen des „Nichtfunktionierens“ wurde schon gestellt. Eventuell verbunden damit ist die Tatsache, dass eine erhebliche Anzahl (Stichprobenprüfung) von „Digitalisierungspositionen“ in verschiedenen politischen Gremien, Beratungsgremien und Verwaltungsorganisationen „fachfremd“ besetzt sind. Kann dies gut gehen? Wohl eher kaum. Wie konnte es zur aktuellen Situation kommen, wie wird es weiter gehen, wie sollte es weiter gehen? Was können wir Informatiker/ -innen tun?

2 Weitere Arbeitsfragen zur Diskussion

1. Sind die derzeitigen Systeme in der Lage, sich selbst hin zu intelligenten Systemen zu transformieren?

2. Brauchen wir aufgrund ggf. vorhandener systemischer Widerstände (weitere) Instanzen und Initiativen, die den Transformationsprozess von Verwaltung und Behörden vorantreiben und steuern?

3. Sollten Informatiker/ -innen, ggf. im Rahmen der GI, Verantwortung übernehmen und zusätzliche, eigene Experten/ -innen-geprägte Initiativen starten, um Intelligente Digitalisierungen auf den verschiedenen Ebenen zu begünstigen/ zu sichern bzw. zu erzwingen? Welche Initiativen sind denkbar und sinnvoll?

4. Welche Möglichkeiten bieten KI-Ansätze und KI-Systeme, Transformationsprozesse zu begünstigen/ zu unterstützen?

5. Sind „verarbeitbare“ Digitale Zwillinge von Unternehmen, Ländern, Staaten realisierbar? Welche technischen und technologischen Anforderungen gibt es für die Erstellung und Verarbeitung von solchen Modellen/ Repräsentationen?

6. Welche weiteren Fragen/ Aspekte könnten zielführend sein?

[5] mindestens einige relevante

Literaturverzeichnis

[BM21] Digitalisierung in Deutschland – Lehren aus der Corona-Krise. Gutachten des Wissenschaftlichen Beirats beim Bundesministerium für Wirtschaft und Energie (BMWi), April 2021.

[Ge19] Geske, U.: Künstliche Intelligenz und Digitalisierung – sind wir auf dem richtigen Weg? In [JHW19].

[JHW19] John, U., Hofstedt, P., Wolf, A. (Hrsg.): Deklarative Ansätze zur Künstlichen Intelligenz – punktuelle Beiträge. Online-Post-Proceedings des 32nd Workshop on (Constraint) Logic Programming, Berlin, 2019.

[Me21] Mertens, P.: Ist Deutschland wirklich ein „digitales Entwicklungsland“ – kann die Institutioneninflation helfen? In Wirtschaftsinformatik & Management 3/2021, Springer Gabler, 2021.

[Re21] Jahresbericht 2021 des Rechnungshofs von Berlin, April 2021.

[Sc18] Scheer, A.-W.: Unternehmung 4.0 – Vom disruptiven Geschäftsmodell zur Automatisierung der Geschäftsprozesse. AWSi Publishing, 2018.

[Si19] Siebel, T.: Digital Transformation: Survive and Thrive in an Era of Mass Extinction. RosettaBooks, New York, 2019.

[WeSe20] Weber, R., Seeberg, P.: KI in der Industrie. Hanser-Verlag, 2020.

Autorinnen- und Autorenverzeichnis

A
Abeck, Sebastian, 207
Abecker, Andreas, 1443, 1505
Ahiagble, Agbodzea Pascal, 1023
Ahouzi, Afrae, 1275
Akyol, Ali, 843
Albayrak, Can Adam, 1593
Alfaro, Federico, 783
Altmann, Jürgen, 723
Alves Junior, Augusto Antonio, 499
Appenzeller, Arno, 533
Arlinghaus, Clarissa Sabrina, 139
Ayon, Safial Islam, 1341

B
Badii, Atta, 1121
Baehr, Julian, 1485
Bahr, Tobias, 1067
Bamberger, Dirk, 1195
Barton, Thomas, 897
Beckmann, Helmut, 259, 871, 1629, 1641
Belaidi, Siwar, 1539, 1551, 1571
Bentele, Manuel, 1417
Berg, Holger, 1485
Berger, Patrick, 307
Berger, Sören-Kristian, 1411
Bernardi, Ansgar, 1473, 1495
Berndl, Mario, 935
Berns, Karsten, 947
Besieda, Andrii, 807
Bhanderi, Jigar, 507
Bhat, Aakash, 479
Biemann, Chris, 1257
Blümm, Mirjam, 1353
Bodammer, Christiane, 693
Boenisch, Franziska, 1227
Bollenbach, Jessica, 393
Boltz, Nicolas, 521
Börner, Andy, 703
Borras, Kerstin, 469
Bors, Michael, 995
Borst, Janos, 271
Böschen, Tula, 283
Bouanani, Oussama, 1227
Breithaupt, Carsten, 1203
Brill, Gerwin, 961
Brink, Henning, 1427
Brinker, Jonas, 347
Brockmann, Carsten, 1195
Brüggen, Marcus, 469
Brüning, Herbert, 795
Bruns, Julian, 1443, 1505
Budde, Matthias, 301, 1517
Buhleier, Laura, 995
Bunse, Mirko, 459
Burkart, Christine, 1353
Büsselmann, Lena, 783

C
Carl, K. Valerie, 195
Carmo Precci Lopes, Alice do, 1485
Challenger, Moharram, 1121
Chircu, Alina, 1181
Chlastak, Maria, 1353
Choi, Jeong-Eun, 651
Chowdhury, Shafayet Hossen, 1341
Christen, Peter, 219
Ciroth, Andreas, 1485
Connor, Patrick L. S., 469

Copurkuyu, Mazlum, 897
Cornaggia-Urrigshardt, Alessia, 81
Czarnecki, Christian, 869, 1203

D
Dang, Michelle, 1093
Danos, Vasilios, 1247
Datta, Kamalika, 1109
Debus, Pascal, 1275
Degeling, Martin, 723
Demus, Christoph, 35
Dhungel, Anna-Katharina, 1305
Diebel-Fischer, Hermann, 755
Diethelm, Ira, 135
Dittert, Nadine, 135
Djeukoua, Roger, 1557
Dodiya, Janki, 131
Dokic, Dusan, 1037
Dorffer, Nina, 1505
Draxler-Weber, Nicole, 1427
Drechsler, Rolf, 1109

E
Eberz-Eder, Daniel, 931, 961
Ebner, Katharina, 717
Egly, Uwe, 1133
Eibl, Maximilian, 703
Eisenhut, Maximilian, 49
Eleks, Marian, 161
Emde, Katharina, 1517
Engbert, Ralf, 1341

F
Faber, Peter, 1085
Fähndrich, Johannes, 71
Fecht, Pascal, 1257
Federer, Marika, 1145
Felser, Jenny, 35
Fischer, Thilo, 1517
Floerecke, Sebastian, 1613
Franken, Jonas, 961
Frey, Tobias, 995
Fritsch, Katharina, 1353
Fuchs-Kittowski, Frank, 935
Fukas, Philipp, 161
Funk, Stefan, 507

G
Gadatsch, Andreas, 1593
Gaede, Frank, 469
Gálfi, Gergely, 1101
Gantner, Patrick, 995
Garrel, Jörg von, 307
Garske, Viktor, 59
Gassen, Eike, 947
Gehrke, Lukas, 219
Gehrke, Melanie, 1203
Gembarski, Paul Christoph, 325
Gerking, Christopher, 521
Gessert, Felix, 1223
Geßner, Hendrik, 1341
Goedicke, Michael, 1317
Götte, Gesa Marie, 1015
Götze, Jana, 823
Griese, Florian, 469
Grimm, Rüdiger, 519
Guiraud, Timothée, 1557
Günnemann, Stephan, 1121
Guo, Yun, 1583

H
Haag, Benedikt, 1593
Hafemeister, Tim, 935
Hagedorn, Tabea, 1485
Hagen, Simon, 325, 1667
Halberstadt, Jantje, 807
Hammer, Sebastian, 1181

Hänel, Albrecht, 1163
Hein, Andreas, 393
Heinisch, Cornelia, 131
Heinke, Florian, 91
Hellmig, Lutz, 755
Herget, Gebhard, 1181
Herzog, Andreas, 1015
Hess, Claudia, 131
Heuschkel, Marie-Luise, 91
Heyde, Markus von der, 1379
Hilbring, Desiree, 1517
Hiller, Axel, 823
Hinz, Oliver, 361
Hinz, Wilko, 823
Hoffmann, Anna, 1081
Höllig, Jacqueline, 301
Holness, Virginia, 693
Holtzblatt, Karen, 151
Homer, Sebastian, 913
Honekamp, Wilfried, 49, 71
Hornung, Gerrit, 519
Höveling, Sven von, 831
Hubertus, Dominik, 361
Hühn, Janis, 1667

I
Ihlenfeldt, Steffen, 1163

J
Jansen, Maike, 1485
Jarocky, Nikita, 81
Jo Pesch, Paulina, 231
Johannsen, Andreas, 885
John, Ulrich, 1681, 1695
Junker, Markus, 1473
Jussen, Ilka, 331, 379

K
Kahle, Reinhard, 1081
Kalvankar, Shreyas, 489
Kamalapur, Snehal, 489
Kammler, Friedemann, 325, 347
Kappler, Karolin Eva, 693, 717
Kasieczka, Gregor, 469
Kaufhold, Marc-André, 995
Keller, Jörg, 99
Keller, Robert, 393
Kellerer, Bernd, 1485
Khalaf, Mahmoud, 1531, 1539
Khaydarov, Valentin, 1061
Kienle, Andrea, 735
Kirdan, Erkin, 1551
Klöpper, Miriam, 1289
Knipprath, Nina, 1359
Knorr, Konstantin, 639
Koch, Christian, 1181, 1203
Kohl, Tobias, 361
Kohler-Bußmeier, Michael, 1565
Kohn, Matthias, 231
Köhne, Sonja, 1289
Kokkarachedu, Sreedhar, 795
Kole, Abhoy, 1109
Kopp-Malek, Tanja, 693
Kortum, Henrik, 283, 361, 1667
Kozsik, Tamás, 1101
Kraft, Angelie, 1257
Kraft, Martin, 807
Krämer, Markus, 119
Krcmar, Helmut, 393
Kreho, Edin, 1557, 1571
Kreinsen, Moritz, 1411
Krempel, Erik, 533
Kültür, Remzi, 1641
Kummer, Janis, 469
Kunau, Gabriele, 735
Kuntke, Franz, 931, 961
Kunz, Sibylle, 131

Kunz, Thomas, 231, 589
Kurth, Frank, 81

L
Labudde, Dirk, 35, 91, 109
Lachowitzer, Jonas, 301
Langer, Fabian, 1247
Lässig, Jörg, 1145, 1153, 1163, 1175
Laurito, Walter, 301
Lehner, Franz, 1613
Leible, Stephan, 1365
Leis, Viktor, 1223
Lemcke, Stefanie, 1399
Lemke, Ole, 723
Lenk, Steve, 1145
Liebelt, Helena, 1085
Losse, Ann-Kathrin, 1203
Lossow, Stefan, 1505
Lowin, Maximilian, 247
Lucke, Ulrike, 1317, 1341, 1399
Ludsteck, Johannes, 1181
Ludzay, Max, 1365

M
Mädler, Jonathan, 1061
Mainzer, Klaus, 427
Majuntke, Verena, 869
Malyshev, Dmitry, 479, 507
Mandl, Alexander, 1133
Markert, Karla, 1275
Markert, Thora, 1247
Marsden, Nicola, 151, 683
Martin, Tobias, 1517
Martinetz, Simone, 683
Marx Gómez, Jorge, 763, 771, 783, 795, 831, 843
Matzner, Matthias, 711
Maurer, Robert, 885
Meier, Pascal, 283
Meinecke, Christofer, 271
Meister, Vera G., 869
Mertzen, Daniela, 1329
Moin, Armin, 1121
Möller, Frederik, 331, 379
Morgenstern, Martin, 71
Moturu, Harish, 763
Müller, David, 639
Müller, Georg, 961
Müller-Birn, Claudia, 745
Müller-Czygan, Günter, 1447
Müllmann, Dirk, 601
Münzberg, Alexander, 1495
Muras, Marco, 831
Müssig, Daniel, 1145, 1163, 1175

N
Naumann, Stefan, 1443
Neubig, Stefan, 393
Neumann, Carolin, 139
Neumann, Frank, 913
Neuroth, Heike, 1329, 1353
Niekler, Andreas, 271
Niemann-Lenz, Julia, 1411
Noack, Andreas, 59

O
Okhovat, Farid, 1195
Orak, Berna, 533
Ortland, Arne, 807

P
Packmohr, Sven, 1427
Pancratz, Nils, 711
Pandit, Hrushikesh, 489
Papke, Alexander, 795
Parwate, Pranav, 489
Päßler, Mario, 823

Pataky, Dominik, 693
Patil, Atharva, 489
Peters, Ludger, 1531, 1539
Petrik, Dimitri, 409
Philipp, Patrick, 301
Pierog, Tanguy, 499
Platzer, Florian, 569
Plociennik, Christiane, 1485
Pourjafarian, Monireh, 1485
Povalej, Roman, 33
Preuß, Svenja, 109
Pröbster, Monika, 683

R
Raabe, Oliver, 521
Rahm, Erhard, 219
Raudonat, Kerstin, 683
Rebholz, Sandra, 131
Rebstadt, Jonas, 161, 283
Reers, Volker, 1153
Reithmair, Christoph, 693
Rettberg, Simon, 1417
Reuter, Christian, 931, 961, 995
Reuter, Willy, 1181
Riedel, Christian, 1341
Ritter, Steffen, 1417
Rodríguez-del-Pino, Juan Carlos, 913
Rößler, Christian, 1417
Rudel, Steffi, 1213
Ruskowski, Martin, 1485
Rustige, Lennart, 469

S
Saleh, Shehab, 1485
Salemi, Simone, 665
Sampathkumar, Pranav, 499
Sänger, Niklas, 207
Saukel, Kevin, 1393
Schebek, Liselotte, 1485
Schering, Johannes, 763, 783
Schieberle, Andreas, 1353
Schindler, Josef, 1539, 1551
Schleper, Janine, 231
Schleper, Peter, 469
Schlömer, Inga, 131
Schlunder, Philipp, 1055
Schmidt, Vanessa, 855
Schmieder, Kai, 683
Schmunk, Stefan, 1353
Schneemann, Carsten, 1329
Schneider, Maja, 219
Schneider, Wolfgang, 961
Schönhofen, Felix, 409
Schoormann, Thorsten, 325
Schreiber, Linda, 237
Schulze, Martin, 935
Schulze, Paul, 935
Schwab, Anne-Kathrin, 807
Schweihoff, Julia Christina, 331, 379
Seegebrecht, Anja, 1341
Seip, Jendrik, 247
Selch, Maximilian, 1163
Selzer, Annika, 159
Sengupta, Indranil, 1109
Seßler, Matthias, 1181
Setz, Tahireh, 651
Sever, Sercan, 729
Simmet, Christian, 935
Simon, Judith, 1257
Sipos, Lars, 745
Smolnik, Stefan, 717
Solsbach, Andreas, 843
Sorge, Christoph, 519
Spiecker genannt Döhmann, Indra, 519
Spiekermann, Daniel, 99

Spranger, Michael, 35
Stachon, Maleen, 331, 379
Steden, Bettina, 771
Steffen, Adrienne, 131
Steffes, Bianca, 665
Stein, Hannah, 1023, 1037
Steinbrink, Enno, 961
Steinebach, Martin, 651
Sterz, Leonie, 521
Stiens, Philipp, 703
Stockhausen, Pablo, 885
Störl, Uta, 1093
Straka, Janine, 1329
Strerath, Paul, 973
Stummer, Sarah, 179
Sturm, Martina, 1195
Stüven, Jan, 783
Suchodoletz, Dirk von, 1417
Sudeikat, Jan, 1565
Sultanow, Eldar, 1181, 1203
Szafarski, Daniel, 871

T
Tarasyuk, Viktoriya, 1447
Tchuegoue Djeukoua, Louis Roger, 1571
Thielert, Bonito Steffen, 1015
Thimm, Heiko, 855
Thoma, Steffen, 301
Thomas, Oliver, 161, 283, 347, 361, 1667
Timm, Ingo J., 159
Töpfer, Heinrich, 831
Trapp, Matthias, 931, 961
Traubinger, Verena, 703
Trieflinger, Stefan, 409
Troost, Christian, 1495
Tschepetzki, Ralf, 1447

U
Ullrich, André, 1203
Ulrich, Ralf, 499
Urban, Alexander, 935
Urbas, Leon, 1061
Urrigshardt, Sebastian, 81
Usbeck, Ricardo, 1257

V
Vaas, Silke, 131
Veit, Maxime, 601
Viedt, Isabell, 1061
Vogel, Daniel, 119
Vogel, Inna, 651
Vogel, Naomi, 507
Vogelgesang, Malte, 1485
Volkamer, Melanie, 601

W
Waedt, Karl, 1531, 1539, 1551, 1557, 1571
Wagner, Gian-Luca, 1629
Waldmann, Ulrich, 231, 589
Wambach, Tim, 617
Wang, Junjie, 1583
Welz, Laslo, 259
Werner, Stephanie, 1353
Widemann, Marc, 551
Wied, Christian, 961
Wiegreffe, Daniel, 271
Wiese, Jürgen, 1447
Wilkinghoff, Kevin, 81
Wingerath, Wolfram, 1223
Winkler, Mareike, 1289
Wintz, Lea, 973
Wittmer, Sandra, 569
Wolf, Armin, 1655
Wolf, Nicole, 693

Wolf, Patrick, 947
Wybrands, Marius, 831

X
Xi, Jian, 35

Z
Zajac, Markus, 1093
Zimborás, Zoltán, 1101
Zinn, Bernd, 1067
Zmija, Andreas, 507
Zorn, Hans-Peter, 1257
Zulauf, Bert, 1359